BIOCHEMICAL CHARACTERIZATION OF MCM4^{CHAOS3} MUTATION USING AN ARCHAEAL MCM AS A MODEL

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Qiuyue Yang

May 2013

ABASTRACT

The minichromosome maintenance (MCM) complex is essential for DNA replication in eukarya and archaea and is thought to function as the replicative helicase. While in eukarya the MCM is a complex of six highly conserved but nonidentical polypeptides (Mcm2-7), most archaeal species contain 6 identical polypeptides. In a N-ethyl-N-nitrosourea (ENU) mutagenesis screen for mice showing chromosome instability, a mutant mouse was identified which harbored a mutation in MCM4 protein referred to as Chaos3 (chromosome aberrations occurring spontaneously 3). MCM4^{Chaos3} mice harbor a single amino acid substitution (F345I) in which a conserved phenylalanine residue in MCM4 protein is replaced by isoleucine. Female mice homozygous for the mutation are susceptible to mammary adenocarcinomas. Similar mutations in the budding yeast homologue of MCM4 resulted in a strong minichromosome loss phenotype. Using the single MCM homologue from the archaeon Methanothermobacter thermautotrophicus (mtMCM), we show that the equivalent mutation (F172I) affects the stability and activity of the helicase. The implications of these observations for the eukaryotic MCM functions are discussed.

Besides the biochemical characterization, I have carried out structural studies of mutant proteins using Small Angle X-ray Scattering (SAXS) to illustrate the structural effect of F172I mutation. Some preliminary results are presented to show that F172I mutation does affect the structural integrity of N-terminal mtMCM protein.

BIOGRAPHICAL SKETCH

The author was born in Urumqi, the capital city of Xinjiang province in northwestern China in 1981. After finishing high school, she entered the School of Life Sciences in University of Science and Technology of China (USTC) in 1999. She finished her four-year study and got her Bachelor degree in Science in 2003. She stayed in the Chinese Academy of Sciences (CAS) Key Laboratory of Structural Biology in USTC as a volunteer for one year, and during which time, she focused on the biochemical and structural studies of two proteins under the guidance of Professors Liwen Niu and Maikun Teng. In 2004, by accepting an offer from Prof. Sol Gruner and Dr Quan Hao, she became a graduate student of MacCHESS and the Cornell Biophysics Program to study biological problems using physical methods. Since then, she has done her research work in the field of Biophysics under the supervision of her thesis advisor Dr. Quan Hao. She passed the "A" examination in the March of 2008 regarding structural study of Mini-chromosome Maintainance protein in *Methanobacterium thermoautotrophicum* (mtMCM) using both X-ray crystallography and Small Angle X-ray Scattering methods. After that, she started her study on biochemical characterizations of MCM4^{chaos3} mutation in mtMCM in collaboration with Prof. Bik Tye and Dr. Zvi Kelman.

To the memory of my grandfather,

Shuyin Yang

ACKNOWLEDGMENTS

I would like to express my deep appreciation to my advisor, Dr. Quan Hao who has been a terrific mentor throughout my Ph.D. study. His great personality and attitude in science have inspired me not only in work but also in life. His warm heart and wisdom always light my life. Without his constant support and guidance, this thesis would not have been possible.

I would like to sincerely thank my supervisor Prof. Sol Gruner for accepting and supporting me as a graduate student of Cornell Biophysics Program, and for his constant guidance throughout my project and exams. I also thank Prof. Richard Cerione for serving on my special committee and providing helpful advices on my project and research equipments to assist me fulfill this thesis.

I would like to especially thank Prof. Bik Tye for her great help throughout this project and my whole Ph.D. study. It is she who introduced this project to me and inspired my enthusiasm to MCM proteins. It is she who brought this collaboration together and supported me all the way to the end. She is not only a great teacher but also a kind friend, who cares about not only my work but also my life track. I feel very fortunate to have met and became a student of such a great professor.

I owe my deepest gratitude to Dr. Zvi Kelman from University of Maryland as a collaborator on this project. He has provided great help throughout this research work, from hosting me in his lab to perform many of the biochemical assays, to guiding me on data analysis, and to assisting me on publication preparation. His passion in science and teaching has a big impression on me, leading me through all the difficulties in the project. I would like to also thank Nozomi Sakakibara, a former student from Dr. Zvi Kelman's lab, who taught me several biochemical assays step-by-step and has been a good collaborator on this project. Without their contribution, this work would not have been accomplished.

I am very grateful to all my colleagues of MacCHESS, in particular, Qingqiu Huang for his experience on protein purification; Qun Liu for his advices and guidance on energy minimization calculations; David Schuller for constant assistance on electronic equipments; Yeyun Zhou for her assistance during my pregnancy and her friendship in life and Kathy Dedrick for all documentary issues. I would like to thank all other members of MacCHESS for their constant scientific, technical, and social assistances to accomplish this work.

I would like to deliver special thanks to my friends in Ithaca, Emily Sherwin and Kevin Clermont. Their friendship always warms my heart. Without their concern and encouragement, I would not go this far.

I would like to take this chance to thank my parents, Xiaoqin Yang and Yi Qiu. You are to me what ATP is to MCM - the driving force. Thank you for your constant love and support. Without your scarifications and huge help, I could not finish this thesis today. I also want to thank my grand parents, Shuyin Yang and Yurui Xiang, for their love and encouragement on my Ph.D study and in life.

I would like to thank my family, my husband Mingbo, for encouraging me to always go the extra mile and strive for excellence; my son Juju, who gives me strength to hang on with his pure love; my daughter Baobao, who makes me feel so lucky and happy everyday. You are to me, what DNA and Mg^{2+} is to MCM - the essence of my life.

This work is supported by the NSF and NIH (RR01646) grants awarded to Quan Hao, a grant from the NSF (MCB-0815646) awarded to Zvi Kelman, by NSF (MBG-0453773) and NIH (GM072557) grants awarded to Bik Tye.

TABLE OF CONTENTS

BIOGRAPHICAL SKETCH....iii

DEDICATION....iv

ACKNOWLEDGEMENTS....v

LIST OF FIGURES....x

LIST OF TABLES....xi

LIST OF ABBREVIATIONS....xii

PREFACE....xiv

CHAPTER 1. INTRODUCTION....1

1.1 DNA replication....1

1.2 Helicases....2

1.3 MCM helicase....2

1.3.1 Eukaryotic MCMs....3

1.3.2 Archaeal MCMs....4

1.4 *Methanobacterium thermoautotrophicum* MCM (mtMCM) protein....5

1.4.1 Biochemical properties of MCM helicases....6

1.4.1.1 Helicase activity....6

1.4.1.2 ATPase activity....6

1.4.1.3 DNA binding activity....7

1.4.2 Structure studies of MCM helicases....7

1.4.2.1 X-ray crystallographic structures....7

1.4.2.2 Electron microscopy (EM) structures....9

1.4.2.3 Comparison of major approaches for protein structure analysis....10

A) X-ray Crystallography.....10
B) Nuclear Magnetic Resonance (NMR).....11
C) Cryo-electron microscopy (Cryo-EM).....11
D) Small-angle X-rays scattering (SAXS).....12
E) Other methods.....12
Reference.....13
CHAPTER 2. BIOCHEMICAL CHARACTERIZATION OF MCM4^{CHAOS3} MUTATION USING AN ARCHAEAL MCM AS A MODEL.....17
2.1 Abstract.....17
2.2 Introduction.....18
2.3 Materials and Methods.....21
2.3.1 Materials.....21
2.3.2 Proteins expression and purification.....21
2.3.3 Gel filtration analysis.....23
2.3.4 Light scattering.....23
2.3.5 Circular dichroism.....24
2.3.6 Helicase assay.....24
a) Preparation of substrates for Helicase Assays.....24
b) DNA Helicase assay.....26
2.3.7 Statistical Model for Doping Assay.....26
2.3.8 ATPase assay.....27
2.3.9 DNA binding assay.....28
2.3.10 Energy minimization.....29
2.3.11 Processivity assay.....29
2.3.12 TrpY protein replacement assay.....30
2.4 Results.....30

2.4.1 Phe172 is needed for efficient helicase activity.....30
2.4.2 Phe172 is required for efficient DNA binding by MCM.....35
2.4.3 Phe172 participates in MCM hexamerization.....38
2.4.4 Structure simulation suggests a thermodynamically less stable model for F172I mutant protein.....41
2.4.5 F172I mutation does not affect processivity activity.....42
2.4.6 F172I mutation does not affect protein displacement activity.....44
2.5 Discussion.....46
Reference.....49
CHAPTER 3. STRUCTURAL STUDIES OF N-TERMINAL F172I MUTANT MTMCM USING SMALL-ANGLE X-RAY SCATTERING (SAXS) METHOD....53
3.1 Introduction.....53
3.2 Materials and Methods.....56
3.2.1 Materials.....56
3.2.2 Small-angle x-ray scattering experiment.....56
3.2.3 SAXS data analysis.....57
3.3 Preliminary Results and Discussion.....57
Reference.....61
APPENDIXA ABSTRACTS FOR ADDITIONAL PUBLISHED WORK.....63

LIST OF FIGURES

Figure 1.1 Crystal structure of N-terminal mtMCM double hexamer.....8

Figure 1.2 Domain structure of N-terminal mtMCM monomer.....9

Figure 1.3 Average images of the different populations found in one mtMCM preparation under EM.....9

Figure 2.1 The location of the Phe 172 residue in the MCM structure.....20

Figure 2.2 mtMCM with Ile substitution for Phe at position 172 exhibits impaired helicase activity.....31

Figure 2.3 mtMCM F172I mutant protein inhibits helicase activity when complexed with the wild-type enzyme.....34

Figure 2.4 Mutations of Phe 172 do not affect DNA binding by mtMCM.....36

Figure 2.5 Mutations of Phe 172 reduce ATPase activity by mtMCM.....37

Figure 2.6 Mutations of Phe 172 affect the oligomeric structure of mtMCM.....40

Figure 2.7 Mutations of Phe172 does not affect the over all secondary structure of mtMCM.....41

Figure 2.8 5-residue simulation models of N-terminal wild-type and mutant proteins.....42

Figure 2.9 The processivity ability of mtMCM mutant proteins were not affected.....44

Figure 2.10 Mutations of Phe 172 do not affect protein replacement activity by mtMCM.....45

Figure 2.11 The role of residue 172 in hexamer formation.....46

Figure 3.1 GASBOR model of mutant N-terminal mtMCM (F172I) from SAXS data at 2mg/ml protein concentration.....58

Figure 3.2 GASBOR model of mutant N-terminal mtMCM (F172I) from SAXS data at protein concentrations ≥ 4mg/ml.....59

LIST OF TABLES

Supplementary Table 1 Sequences of substrates used in processivity and TrpY displacement assays.....52

LIST OF ABBREVIATIONS

AAA+	ATPases associated with various cellular activities
bp	Base pair
BSA	Bovine serum albumin
CD	Circular dichroism
Cdc6	Cell division cycle 6
Chaos3	Chromosome aberrations occurring spontaneously 3
dsDNA	Doublestranded DNA
EM	Electron microscope
H2ins	Helix 2 insertion
HTH	Helix turn helix
IR	Inverted repeat
ITC	Isothermal titration calorimetry
kDa	Kilodalton
MCM	Minichromosome maintenance
Mka	Methanococcus kandleri
Mt	Methanothermobacter thermautotrophicus
OB	Oligonucleotide/oligosaccharide binding
OBP	Origin binding protein
ORB	Origin recognition box
ORC	Origin reognition complex
ORP	Origin recognition protein
PS1BH	Presensor I b hairpin
PSII	Presensor II insertion
RF	Arginine finger
SAXS	Small Angle X-ray Scattering

SDS	Sodium dodecyl sulfate
SDSPA	GE SDSPolyacrylamide gel electrophoresis
ssDNA	Singlestranded DNA
Sso	Sulfolobus solfataricus
Tm	Melting temperature
WA	Walker A
WB	Walker B

PREFACE

To maintain the integrity of the genome, cell needs to precisely regulate a series of important molecular events, such as DNA replication, chromosomal segregation and surveillance and repair of damaged DNA. One of the key players in replication is the replicative helicase. The minichromosome maintenance (MCM) complex also known as the replicative helicase is involved in DNA replication initiation and elongation in eukaryotes and archaea, the MCM proteins were first identified as protein required for mini-chromosome maintenance in *Saccharomyces cerevisiae* by Bik-Kwoon Tye's laboratory in the early 1980s. MCM is also the target of various checkpoint pathways, such as the S-phase entry and S-phase arrest checkpoints. Both the loading and activation of MCM helicase are strictly regulated and are coupled to cell growth cycles. Defects in MCM function has been linked to genomic instability and a variety of carcinomas (Chapter I).

During a N-ethyl-N-nitrosourea (ENU) mutagenesis screen for mice showing chromosome instability, a mutant mouse was identified which harbored a mutation in MCM4 protein referred to as Chaos3 (chromosome aberrations occurring spontaneously 3). The mutation in $Mcm4^{Chaos3}$ is a substitution of a conserved Phe residue by Ile. It was found that females that are homozygous for the mutation are susceptible to mammary adenocarcinomas with a mean latency of 12 months, whereas homozygous males did not show any significant increase in tumor incidence. It was also found that the mutation affects the stability of the MCM4. Similar mutations in the budding yeast homologue of MCM4 resulted in a strong minichromosome loss phenotype. The yeast mutant bearing the equivalent allele recapitulates many of the mouse mutant phenotype including chromosome instability, G2/M delay and a predisposition of the mutant to acquire improved

proliferation rate. But the biochemical mechanism of how this mutation affects the emzyme is not clear.

The Phe residue, mutated in MCM4^{Chaos3}, is not only conserved in the eukaryotic MCM4 proteins but also in the archaeal MCM protein. Thus, the MCM homologue from the *archaeon Methanothermobacter thermautotrophicus* (mtMCM) was used to study the effect of MCM4^{Chao3} mutation on MCM functions. All archaeal MCM proteins studied to date are from organisms with a single MCM gene in their genome thus is considered as a simple model for understanding the MCM function and structure relationship. Biochemical characterization of both full-length and N-terminal mtMCM proteins harboring the equivalent mutation of MCM4^{Chao3} have been performed to reveal the effect of mutation F172I on the enzyme (Chapter II).

Structural effects of equivalent MCM4Chao3 mutation (F172I) on N-terminal mtMCM have also been investigated. Until recently there is no high resolution structure of the full length mtMCM protein available. Although Xraycrystallographic structures of the N-terminal portion of mtMCM has been solved, my efforts to crystallize N-terminal F172I mutant protein failed. Thus we used structure simulation and Small Angle X-ray Scattering (SAXS) to examine the structural effects of F172I mutation on N-terminal mtMCM protein. Preliminary results are presented and discussed as a supplementary to the biochemical characterization in Chapter III.

CHAPTER 1

INTRODUCTION

1.1 DNA replication

To maintain the integrity of a genome, the cell needs to precisely regulate a series of important genetic events, such as DNA replication, chromosomal division and recognition and reparation of damaged DNA. Defects in any of these processes may cause genomic instability, potentially leading to cancer or death (1).

DNA replication is one of the most essential events in the cell that involves dozens of proteins and enzymes to ensure that genetic information is duplicated in an accurate and timely manner. DNA replication is functionally and often structurally conserved in all life forms and is divided into three main stages: initiation, elongation and termination (2). During the S phase of the cell cycle, DNA replication is initiated at particular points within the genome, known as replication "origins", which contain DNA sequences recognized by replication initiator proteins. Initiator proteins recruit other origin binding proteins (OBP) to separate the DNA strands at the origin, forming a bubble. The replicative helicase is recruited to unwind the double stranded (ds) DNA in a bidirectional manner and the ssDNA binding proteins (SSB) bind to ssDNA to prevent it from reannealing or breaking down. Once strands are separated, RNA primers are synthesized on the template strands and DNA polymerase extends these to create newly synthesized DNA. As DNA synthesis continues, the parental DNA strands continue to unwind on each side of the bubble, forming replication forks. Because bacteria have a single origin of replication in their circular chromosome, the termination of replication occurs when the two progressing replication forks meet each other in the circular chromosome. In contrast, eukaryotes initiate DNA replication at multiple origins in their linear chromosome, so replication forks meet and terminate at

many points in the chromosome, but it is unclear that how the termination of replication ends (3,4).

One of the key events in the initiation of replication is the assembly of helicase at the origin. The minichromosome maintenance (MCM) protein is known as the replicative helicase in eukaryotes and archaea, which is an essential factor in DNA replication initiation and elongation (5).

1.2 Helicases

Helicases, a group of enzymes that catalyze the unwinding of duplex DNA, are required for accessing single strand DNA (ssDNA) template and separating template DNA strands on replication forks. They use the energy generated in enzymatic nucleoside triphosphate (NTP) hydrolysis to translocate along one strand of the DNA duplex, thus displacing the complementary strand from the duplex (6). DNA helicases work in association with other proteins as part of a complex machine in DNA replication, repair, and recombination. However, in vitro, most helicases by themselves are able to catalyze the strand separation or unwinding process with the presence of suitable DNA or RNA substrate (7). Helicases can unwind substrate in different unwinding directions: $5' \rightarrow 3'$ (as dnaB helicase) or $3' \rightarrow 5'$ (as MCM helicase). Though most of helicases unwind either DNA or RNA, some can unwind both (7).

The integrity of genetic information depends on efficient helicase reactions, therefore, it is not surprising that defects in helicases may lead to disorders in the cell. Most of these disorders are characterized by genetic instability caused by defects in DNA repair, recombination, or replication processes (7).

1.3 MCM helicase

The minichromosome maintenance (MCM) protein was first identified as protein required for mini-chromosome maintenance in *Saccharomyces cerevisiae* by Bik-Kwoon Tye's laboratory in the early 1980s. MCM proteins are known as the replicative helicase involved in DNA replication initiation and elongation in eukaryotes and archaea. They are also implicated in many other chromosome transactions including transcription, chromatin remodeling, and genome stability (1).

The MCM proteins are distinct members of AAA ATPase family, which has many cellular functions. AAA ATPases generally form large ATP-dependent complexes, often heterohexamers. The MCMs contain a characteristic version of the ATPase domain, the "MCM box", which spans 200 residues. The MCM box includes two ATPase consensus motifs, Walker A and Walker B. The Walker A motif contains the P-loop of the active site. The Walker B motif contains the sequence IDEFDKM, which is conserved in all MCM proteins and defines the MCM family, and is thought to contribute to ATP hydrolysis. Another short motif, SRFD, called the "arginine finger", usually occurs about 70 residues after the Walker B element. All MCMs contain these sequences. All eukaryotic MCMs have zinc-binding motifs, and studies show that the zinc-binding motifs in several MCMs are required for viability, and may also contribute to complex assembly and ATPase activity (1,5).

1.3.1 Eukaryotic MCMs

In eukarya, the MCM complex consists of six essential proteins (Mcm2−7) that each contains a conserved C-terminal helicase domain and divergent N-terminal sequence. Mcm2-7 proteins form a heterohexameric complex, which is a key component of the pre-replication complex (pre-RC) that assembles on replication origins before S phase. The assembly of the pre-RC, involving several conserved initiation proteins like the ORC (origin recognition complex), Cdc6, Cdt1, and the

Mcm2-7 complex, initiates DNA replication which is strictly controlled so that chromosomal DNA is duplicated only once per cell cycle (5,8).

In vivo experiments indicate that all six proteins are required for the initiation and elongation steps of replication (9). In addition to forming a heterohexamer, in vivo and in vitro studies have revealed the existence of several MCM subcomplexes (10). Biochemical studies have shown that a dimeric complex of the Mcm4,6,7 heterotrimer possesses weak 3'→5' DNA helicase activity, can translocate on single and double stranded DNA, can bind DNA and RNA, and has the ability to unwind DNA-RNA duplexes while translocating on the DNA strand. The interaction of Mcm4,6,7 complex with either MCM2 or the MCM3,5 dimeric complex inhibits the helicase activity, suggesting regulatory function of these Mcm proteins (11,12). A recent study detects helicase activity of the Mcm2-7 complex in vitro, although it is strongly anion dependent. It was also suggested that that Mcm2,5 plays a role in regulation of Mcm4,6,7, possibly as a gate of the ring opening in order to be loaded onto the DNA in the Mcm2-7 complex (13).

The archaeal MCM and all eukaryotic MCMs share a well conserved C-terminal domain, which contains the helicase domain, similar to other hexameric helicases functioning in origin melting and fork DNA unwinding. The N-terminal domain, less conserved among all MCMs, binds DNA and is required for multimerization (5).

1.3.2 Archaeal MCMs

There are three domains of life, bacteria, eukarya and archaea (14). Although archaea share similarity to bacteria in terms of cell structure and metabolism due to their prokaryotic nature, some archaeal proteins involved in replication, transcription, and translation share more similarities with those from eukaryotes (15).

MCM homologs have also been identified in archaea, and at least one MCM homologue has been identified in all archaea sequenced to date (16). Based on structural and biochemical features it is thought that the archaeal MCM proteins function as the archaeal replicative helicase. Several archaeal MCM helicases have been studied, and their biochemical properties have been described. The helicases from most archaea form hexamers in solution, have an ATP-dependent 3'→5' helicase activity, bind and translocate along ssDNA (single-stranded DNA) and dsDNA (double-stranded DNA), can unwind DNA–RNA hybrids while moving along the DNA strand, and can displace proteins from DNA (17).

Currently, the following archaeal MCM X-ray structures have been determined: the N-terminal structures for mthMCM (18) and ssoMCM (19), a near full-length ssoMCM structure (20), and an MCM homolog from another archaeon, M. kandleri (mkaMCM2) (21) (reviewed in Ref. 17). These archaeal MCM structures have provided valuable information for understanding the enzymatic mechanism of MCM helicase.

1.4 *Methanobacterium thermoautotrophicum* MCM helicase

Methanothermobacter thermautotrophicus is a methanogenic archaeabacteria which grows and gains energy through the conversion of hydrogen and carbon dioxide to methane with an optimal growth temperature between 62 ~ 65ºC (22). The *M. thermautotrophicus* genome contains one replication origin, one eukaryotic MCM helicase homologue, and two homologues of the putative helicase loader and OBP, Cdc6 (23). The *M. thermoautotrophicus* MCM homolog (MtMCM, ~76 kDa), similar to the eukaryotic MCM4 (1,5), has been shown to have helicase activity in vitro (16). *M. thermautotrophicum* MCM is 34% identical and 51% similar in sequence to the MCM2-7 protein family with the strongest homology concentrated in three conserved

motifs, two of which are the Walker A and B nucleotide binding sequence (24). Thus this simpler system is a good model for eukaryotic MCM complex and to establish its role in replication.

1.4.1 Biochemical properties

M. thermoautotrophicus MCM helicase has similar enzymatic activities as eukaryotic MCM complex. It binds DNA, has DNA-stimulated ATPase activity, and unwinds duplex DNA in the 3'→5' direction in an ATP-dependent reaction (17). All these features make mtMCM protein a simpler experimental model for studying helicase function in DNA replication in both archaeal and eukaryotic systems.

1.4.1.1 Helicase activity

All archaeal MCM proteins unwind substrates in the direction of 3'→5'. Substrates with 3' overhang available are required for helicase assay and better activity is detected when a 5' overhang is also present [25-28]. MtMCM protein was also shown to prefer the forked like substrate (both 3' and 5'- overhang dsDNA) over a flat substrate (only 3'-overhang dsDNA) (27). As it translocates along the DNA, the MtMCM helicase can displace proteins such as transcription factor and histones bound on DNA from the substrate (29). MtMCM can also displace DNA-RNA duplex substrate while translocating along the DNA, suggesting multiple roles besides that of a helicase in DNA replication (29).

1.4.1.2 ATPase activity

Helicases are motor proteins that couple the hydrolysis of nucleoside triphosphate (NTPase) to nucleic acid unwinding (30). All archaeal MCM helicases have ATPase activity, some of which can be stimulated in the presence of DNA. In

MtMCM protein, ATPase activity is strongly dependent on the presence of either ss or dsDNA (31) and ATP binding enhances DNA binding by the helicase (25).

1.4.1.3 DNA binding

Archaeal MCM helicases must interact with DNA to have helicase activity. Binding constants (Kd) of archaeal MCM to different DNA substrates have been reported to vary from 10 to 300 nM (32). ATP stimulates the DNA binding ability of MtMCM protein, although the stimulatory effect of ATP on DNA binding can only be detected when the reaction is incubated at 50-60ºC. No change in DNA binding could be detected in the presence of ATP when the DNA binding reactions were incubated at 22ºC (26,32).

1.4.2 Structural studies of *M. thermautotrophicus* MCM helicase

The *M. thermoautotrophicus* MCM can be divided into three parts. An N-terminal region, a middle region, which contains the helicase catalytic domains, and a C-terminal region which was suggested to fold into a helix-turn-helix motif (20,21). High-resolution structural information is not yet available for the full-length MtMCM protein, but an X-ray crystallographic structure of the N-terminal portion of MtMCM has been reported. EM studies have shed light on the overall MtMCM structure as well as its unique properties.

1.4.2.1 X-ray crystallographic structures

A 3.1 Å crystal structure of the N-terminal domain (N-mtMCM) reveals a unique double-hexamer architecture: a dumbbell, slim in the center with large bulges at both ends (18). The dodecamer has an unusually long (118 Å) and highly positively

charged central channel that binds dsDNA. The wall of the channel is made of β-strands, while the bulges around the outside are α-helices.

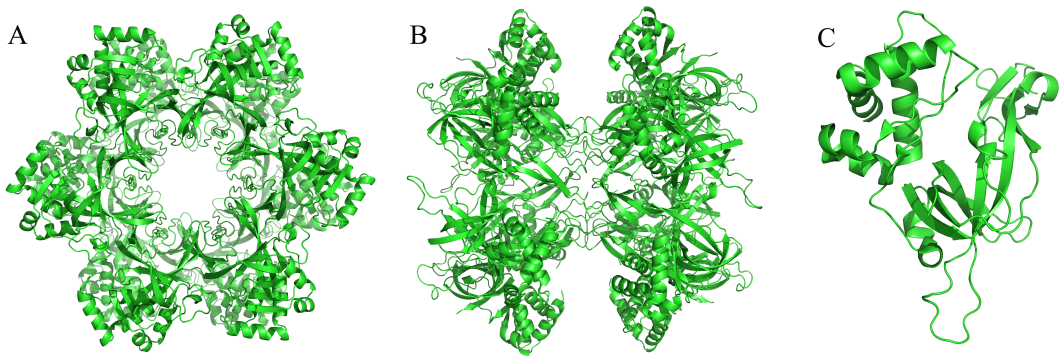


Figure 1.1 Crystal structure of N-terminal mtMCM double hexamer. A. Top view of double-hexamer structure. B. Side view of double-hexamer structure. C. A monomer structure of N-mtMCM.

Each unit has a three-domain structure (16). Domain A, at the N terminus, is mostly α-helical. Domain B has three β-strands and contains a zinc-finger motif shown to be needed for DNA binding. Domain C contains five β-strands that form an oligonucleotide/oligosaccharide binding (OB) fold.

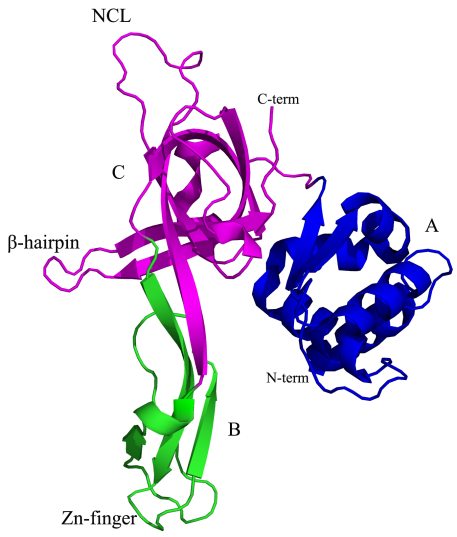


Figure 1.2 Domain structure of N-terminal mtMCM monomer. Domain organization and the main structural motifs are shown. The N-terminal part is divided into three domains, A (Blue), B (Green) and C (Magentas). NCL, N-terminal communication loop; C-term, C-terminal end; N-term, N-terminal end.

This domain connects the N-terminal portion of the enzyme to the C-terminal catalytic region. Domain C contains a zinc-finger motif shown to be essential for ssDNA and dsDNA binding and a conserved loop, N-terminal communication loop (NCL), which plays a role in communication between the N-terminal DNA binding region and the catalytic part. Domain C was also shown to play a major role in MCM multimerization.

1.4.2.2 Electron microscopy (EM) structures

Low-resolution electron micrograph reconstruction studies have been performed on the full-length MCM protein from the archaeon M. thermautotrophicus.

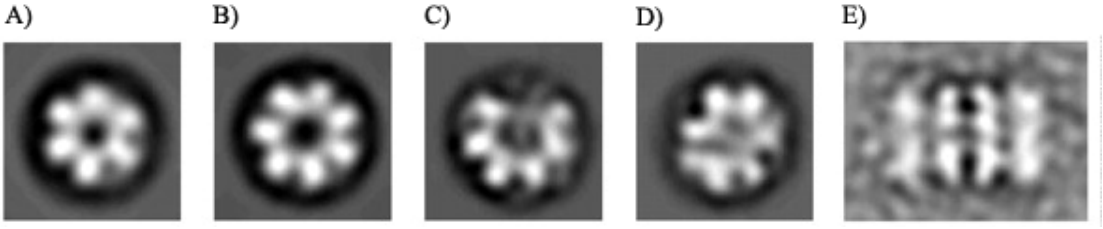


Figure 1.3 Average images of the different populations found in one mtMCM preparation under EM. This figure is cited from "Ref. 33 Figure 2". These images were obtain under solution containing 250mM NaCl, 2mM AMP-PNP, and 5mM NTP after 1 hour incubation at 37 °C. A) 6-fold ring view; B) 7-fold ring view; C) open ring view; D) 2-fold titled ring view; E) side view.

The first image that MtMCM obtained by EM revealed a cylindrical structure

showing that it exists as a double hexamer (26). MtMCM was also found to assemble into a heptameric ring by EM, together with many larger aggregates in the image, consistent with side views of 'stacks' of rings (24). A recent EM study of full-length MtMCM revealed several oligomeric forms (33). Two different types of morphology were found in series of solutions. One is ring-shaped objects, top views of either a single or a double ring. The other is rectangle-shaped particles with four density bands, consistent with side views of a double ring. Analysis of ring views revealed a mixture of 6-fold and 7-fold symmetry forms. Open ring forms were also found in some conditions. In the double ring the 6-fold arrangement was preferred, forming the double hexamer structure.

These studies suggest that full-length MtMCM protein can adopt multiple forms in solution: hexamers, heptamers, dodecamers, filaments and open rings. These structures may be dependent on the treatment of the samples prepared for the EM studies. Nonetheless, the flexibility to form various oligomers may be a mechanism of regulation.

1.4.2.3 Comparison of major approaches for protein structure analysis

A) X-ray Crystallography

Around 90% of the protein structures available in the Protein Data Bank were determined by X-ray crystallography. This technique measures the 3D density distribution of electrons in the protein (in the crystallized state) and thereby infers the 3D coordinates of all the atoms to be determined. Comparing to other methods, X-ray crystallography can produce the highest resolution data (up to 1 Å) revealing fine details of atomic structures. It requires high quality protein crystals to perform the experiment but not all proteins can be crystallized, particularly in the case

of large protein complexes. As the structure information is obtained under crystallized state, flexible portions of protein molecules are sometimes not seen with X-ray crystallography and the protein structure may be influenced by crystal packing forces, which arguably can lead to undesirable conformational changes (34).

B) Nuclear Magnetic Resonance (NMR)

Roughly 9% of the known protein structures have been obtained by the Nuclear Magnetic Resonance (NMR) technique, which can also be used to determine secondary structure. It has relatively high resolution (2-3 Å) and it examines proteins in solution, which is close to physiological environment than crystals. But NMR has a strict limitation on protein size, hardly applicable for protein whose molecular weight exceeding 30 kDa (34).

C) Cryo-electron microscopy (Cryo-EM)

Cryo-electron microscopy (Cryo-EM) has recently become a method of determining 3D protein structure to a moderate resolution (~10Å). This technique is a valuable resource for studying very large protein complexes such as virus coat proteins and amyloid fibers. Compared to the above two methods, Cryo-EM uses low amount of protein and gives direct visualization of particle shape and symmetry. Cryo-EM has lower resolution than X-ray crystallography or NMR and it also has size limitation, hardly applicable for protein molecules whose molecular weights are less than 200 kDa (34).

D) Small-angle X-rays scattering (SAXS)

Small-angle X-rays scattering (SAXS) is another fundamental tool in the study of biological macromolecules. The major advantage of the method lies in its ability to provide structural information about partially or completely disordered systems. SAXS allows one to study native particles in nearly physiological solutions and to analyze structural changes in response to variations in external conditions thus yielding valuable information on the relationship between these structural changes and the operation of these systems. Despite its low resolution limitation (10-20Å), the method is a general tool for studying the structure, kinetics and interactions of biological macromolecules in solution. SAXS is applicable to macromolecules with very different sizes (few kDa to hundreds MDa). Moreover, for complexes with molecular mass of a few hundred kDa, too large for NMR and too small for cryo-EM, SAS remains effectively the only method to obtain their shape in solution. The biggest limitation of this method is that it requires additional information to resolve ambiguity in model building, so it usually needs to work together with other major methods (34).

E) Other methods

Static and dynamic light scattering and ultracentrifugation methods are also used to gain 3D information of biological macromolecules. These methods are simple to setup, non-destructive to samples and requires low amount of proteins. But they only yield overall parameters of the protein molecules in solution, and thus are used to provide supportive information for other high-resolution methods (34).

REFERENCE

1. Forsburg, S. L. (2004) *Microbiol Mol Biol Rev* **68**, 109-131
2. Kelman, Z. and White, M.F. (2005) Archaeal DNA replication and repair. *Curr Opin Microbiol*. December; 8(6): 669–676.
3. Stryer, L. (1995) *Biochemistry* 4th ed. W.H. Freeman and Company, New Y ork.
4. Weaver, R. (2002) *Molecular Biology*. 2nd ed. McGraw-Hill.
5. Tye, B.K. (1999) MCM proteins in DAN replication. *Annual review of biochemistry*, **68**, 649-686.
6. Kelman, Z. and Hurwitz, J. (2003) Structural lessons in DNA replication fro m the third domain of life. *Nat Struct Biol*, **10**, 148-150.
7. P*atel,* S. S. and Picha*,* K. M. (2000) Structure and Mechanism of Hexameric Helicases. *Ann. Rev. Biochem*. **69**, 651-97.
8. M. Lei, I. H. Cheng, L. A. Roberts, M. A. McAlear, and *B. K. (2001) Tye Nucleic Acids Res.*, December 1, **29**(23): 4834 – 4842**.**
9. K. Labib, J. A. Tercero, and J. F. Diffley 2000 Uninterrupted MCM2-7 function required for DNA replication fork progression. *Science.* June 2; **288**(5471): 1643–1647.
10. Tye, B.K. and Sawyer, S.L. (2000) The Hexameric Eukaryotic MCM Helicase: Building Symmetry from Nonidentical Parts. *The Journal of biological chemistry*, **275**, 34833-34836.
11. You, Z., Y. Komamura, and Y. Ishimi. (1999) Biochemical analysis of the intrinsic Mcm4-Mcm6-Mcm7 DNA helicase activity. *Mol. Cell. Biol.* **19**, 8003–8015.
12. Lee, J. K., and J. Hurwitz. (2000) Isolation and characterization of various

complexes of the minichromosome maintenance proteins of Schizosaccharomyces pombe. *J. Biol. Chem.* **275**, 18871–18878.

13. *Bochman, M.L. & Schwacha, A. (2008)* The Mcm2-7 complex has in vitro helicase activity. *Molecular Cell* **31**, 287–293
14. Woese, C.R., Kandler, O. and Wheelis, M.L. (1990) Towards a natural systemof organisms:proposal for the domains Archaea, Bacteria, and Eucary a. *Proceedings of the National Academy of Sciences of the United States of America*, 87, 4576- 4579.
15. Olsen, G.J. and Woese, C.R. (1997) Archaeal genomics: an overview. *Cell*, **89**, 991-994.
16. Kelman, L.M. and Kelman, Z. (2003). Archaea: an archetype for replication initiation studies? *Mol. Microbiol.* **48**, 605-615.
17. *Sakakibara N, Kelman LM, Kelman Z.* (*2009*) Unwinding the structure and function of the archaeal MCM helicase. *Mol Microbiol **72**(2), 286-296.*
18. Fletcher, R.J., Bishop, B.E., Leon, R.P., Sclafani, R.A., Ogata, C.M. and Chen, X.S. (2003) The structure and Function of MCM from archaeal *M. thermoautotrophicum*. *Nature Struct Biol*, **10**, 160-167.
19. Liu, W., Pucci, B.,Rossi, M., Pisani, F.M. and Ladenstein, R. (2008) Structural analysis of the *Sulfolobus solfataricus* MCM protein N-terminal domain. *Nucleic acids research*, **36**, 3235-3243.
20. Brewster, A.S., Wang, G., Yu, X., Greenleaf, W.B., Carazo, J.M., Tjajadi, M., et al. (2008) Crystal structure of a near full-length archaeal MCM: functional insights for an AAA+ hexameric helicase. *Proc Natl Acad Sci USA* **105**, 20191–20196.
21. Bae, B., Chen, Y.H., Costa, A., Onesti, S., Brunzelle, J.S., Lin, Y., et al. (2009) Insights into the architecture of the replicative helicase from the

structure of an archaeal MCM homolog. *Structure* **17**, 211–222.

22. Zeikus, J.G. and Wolfe, R.S. (1972) Methanobacterium thermoautotrophicus sp.n., an anaerobic, autotrophic, extreme thermophile. J Bacteriol, 109, 707-715.
23. Grabowski, B. and Kelman, Z. (2003) Archaeal DNA replication: Eukaryal Proteins in a Bacterial Context. *Annu Rev Microbiol*, **57**, 487-516.
24. Yu, X., VanLoock, M.S., Poplawski, A., Kelman, Z., Xiang, T., Tye, B.K. and Egelman, E.H. (2002) The *Methanobacterium thermoautotrophicum* MCM protein can form heptameric rings. *EMBO Rep*, **3**, 792-797.
25. Kelman, Z., Lee, J.K. and Hurwitz, J. (1999) The single minichromosome maintenance protein of *Methanobacterium thermoautotrophicum* ΔH contains DNA helicase activity. *Proceedings of the National Academy of Sciences ofthe United States of America*, **96**, 14783-14788.
26. Chong,J.P., Hayashi, M.K., Simon, M.N., Xu, R.M. andStillman, B. (2000) A double-hexamer archaeal minichromosome maintenance protein is an ATP-dependent DNA helicase. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 1530-1535.
27. Shin, J.H., Grabowski, B., Kasiviswanathan, R., Bell, S.D. and Kelman, Z. (2003) Regulation of minichromosome maintenance helicase activity by Cdc6. *The Journal of biological chemistry*, **278**, 38059-38067. Barry, E.R., 2007;
28. Rothenberg, E., Trakselis, M.A., Bell, S.D., and Ha, T. (2007) MCM fork substrate specificity involves dynamic interaction with the 5′-tail. *J Biol Chem* **282**, 34229–34234.

29. Shin, J.H., Santangelo, T.J., Xie, Y., Reeve, J.N. and Kelman, Z. (2007) Archaeal minichromosome maintenance (MCM) helicase can unwind DNA bound by archaeal histones and transcription factors. *The Journal of biological chemistry* **282**, 4908-4915.
30. Patel, S. S. and Picha, K. M. (2000) Structure and function of hexameric helicases. *Annual Review of Biochemistry* **69**, 651-697.
31. Jenkinson, E.R. and Chong, J.P. (2006) Minichromosome maintenance helic ase activity is controlled by N- and C-terminal motifs and requires the ATPase domain helix-2 insert. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 7613-7618.
32. Sakakibara, N., Kasiviswanathan, R., Melamud, E., Han, M., Schwarz, F.P., and Kelman, Z. (2008) Coupling of DNA binding and helicase activity is mediated by a conserved loop in the MCM protein. *Nucleic Acids Res* **36**, 1309–1320.
33. Gomez-Llorente, Y., Fletcher, R.J., Chen, X.S., Carazo,J.M. and Martin, C.S . (2005) Polymorphism and Double Hexamer Structure in the Archaeal Minichromosome Maintenance (MCM) Helicase from Methanobacterium thermoautotrophicum. The Journal of biological chemistr y, 280, 40909-40915.
34. Svergun, D.I., and Koch MH. (2003) Small-angle scattering studies of biological macromolecules in solution. *Rep. Prog. Phys.* **66**, 1735-1782.

CHAPTER 2

BIOCHEMICAL CHARACTERIZATION OF MCM4^{CHAOS3} MUTATION USING AN ARCHAEAL MCM AS A MODEL

2.1 Abstract

The minichromosome maintenance (MCM) complex is essential for DNA replication in eukarya and archaea and is thought to function as the replicative helicase. While in eukarya the MCM is a complex of six polypeptides (Mcm2-7), most archaeal species contain a single homologue. MCM4^{Chaos3} mice harbor a single amino acid substitution in which a conserved phenylalanine residue in MCM4 protein is replaced by isoleucine. Female mice homozygous for the mutation are susceptible to mammary adenocarcinomas. Using the single MCM homologue from the archaeon Methanothermobacter thermautotrophicus, we show that the equivalent mutation (F172I) affects the stability and activity of the helicase. The implications of these observations for the eukaryotic MCM functions are discussed.

2.2 Introduction

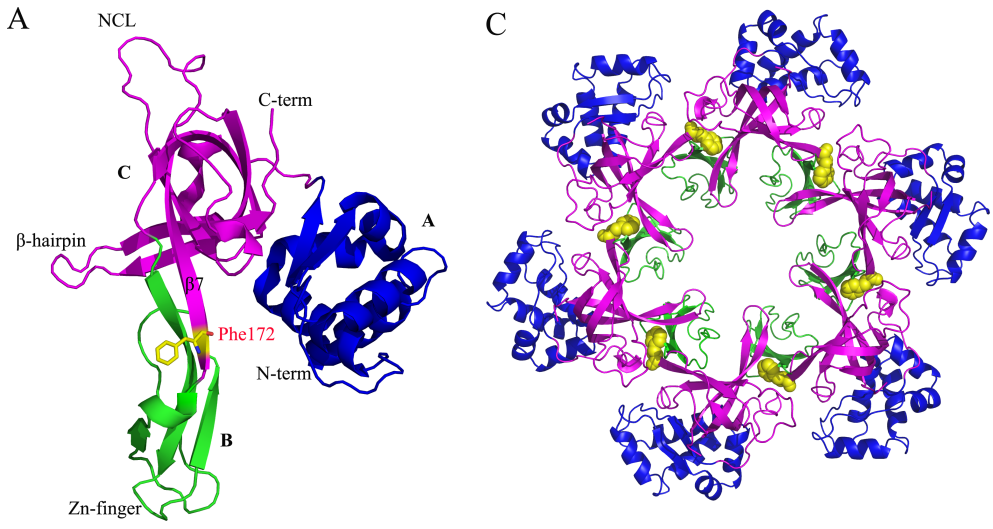
The integrity of genomic information depends on the precise regulation of DNA replication, a process which involves a complicated network of proteins. The helicase is one of the enzymes that play an essential role in the process. The mini-chromosome maintenance (MCM) is thought to function as the replicative helicase in eukarya. MCM is a family of six conserved proteins, Mcm2-7, that forms a heterohexameric complex. It was shown that all six proteins are essential for the initiation and elongation phases of DNA replication (reviewed in: (1-4)).

In a N-ethyl-N-nitrosourea (ENU) mutagenesis screen for mice showing chromosome instability, a mutant mouse was identified which harbored a mutation in MCM4 protein referred to as Chaos3 (chromosome aberrations occurring spontaneously 3) (5,6). The mutation in $Mcm4^{Chaos3}$ is a substitution of a conserved Phe residue by Ile (5). It was found that females that are homozygous for the mutation are susceptible to mammary adenocarcinomas with a mean latency of 12 months, whereas homozygous males did not show any significant increase in tumor incidence (5,6). It was also found that the mutation affects the stability of the MCM4 in vivo as reduced amount of proteins was observed in cell harboring the $MCM4^{Chaos3}$ mutation (5). Similar mutations in the budding yeast homologue of MCM4 resulted in a strong minichromosome loss phenotype (5). A yeast mutant bearing the equivalent allele was recently characterized. The yeast mutant recapitulates many of the mouse mutant phenotype including chromosome instability, G2/M delay and a predisposition of the mutant to acquire improved proliferation rate (8).

MCM has also been suggested to function as the replicative helicase in archaea (reviewed in: (9-12)). All archaeal MCM proteins studied to date are from organisms with a single gene encoding for MCM in their genome. Biochemical studies with several archaeal MCM proteins revealed biochemical properties which are similar to those of the eukaryotic MCM complex. They include ATP-dependent 3′→5′ helicase activity, binding and translocating along single-stranded (ss) and double-stranded (ds) DNA and unwinding of DNA–RNA hybrids while moving along the DNA strand and to displace proteins from DNA (reviewed in: (10,12)).

The archaeal MCM helicases can be divided into three parts; N-terminal region needed for protein multimerization, a middle part which contains the

helicase catalytic domains, and a C-terminal region which was suggested to fold into a helix-turn-helix motif (13,14). High-resolution three-dimensional structures of the N-terminal portion of the MCM proteins from Methanothermobacter thermautotrophicus (15) and Sulfolobus solfataricus (16) revealed a three domain structures (Fig. 2.1A). Domain A, at the N terminus, is mostly a-helical. Domain B has three β-strands and contains a zinc-finger motif shown to be needed for DNA binding (17-19). Domain C contains five b-strands that form an oligonucleotide/oligosaccharide binding (OB) fold. This domain connects the N-terminal portion of the enzyme to the C-terminal catalytic region. Domain C contains a β-finger motif shown to be essential for ssDNA and dsDNA binding (15,18) and a conserved loop, the N-terminal communication loop (NCL), which plays a role in communication between the N-terminal DNA binding region and the catalytic portion (20,21). Domain C was also shown to play a major role in MCM multimerization (22).



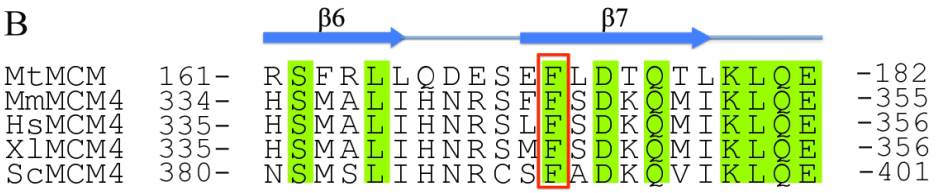


Figure 2.1 The location of the Phe 172 residue in the MCM structure. A. Phenylalanine 172 of mtMCM is conserved in the eukaryotic MCM4 proteins. Sequence alignment of the region of mtMCM containing F172 with MCM4 proteins from *Mus musculus*, MmMCM4; *Homo sapiens*, HsMCM4; *Xenopus laevis*, XlMCM4 and *Saccharomyces cerevisiae*, ScMCM4. Residues that are conserved in all sequences are shown in green and Phe 172 of mtMCM is boxed in red. B. The structure of the mtMCM N-terminal region. Domain A is shown in blue, domain B in green and domain C in magenta. Phe172 is labeled in yellow. Several motifs are also marked. C. The hexameric structure of the mtMCM N-terminal region. Colors are as in panel B.

The Phe residue, which is mutated in MCM4^{Chaos3}, is not only conserved in the eukaryotic MCM4 proteins but also in the archaeal MCM protein (Fig. 2.1B). Thus, the MCM homologue from the archaeon M. thermautotrophicus (mtMCM) was chosen to study the effect of MCM4^{Chao3} mutation on MCM functions. It has been found that the complex formed using the mutant MCM is less stable in comparison to the wild-type enzyme and has reduced helicase activity. The data presented here provide an explanation for the observations made with the mutant mice.

2.3 Materials and Methods

2.3.1 Materials

ATP and [γ-^{32}P]ATP were purchased from GE Healthcare, and oligonucleotides were synthesized by the CARB DNA synthesis facility. The ROX labeled DNA was purchased from Integrated DNA technologies. The wild-type full-length mtMCM construct was original from Prof. Zvi

Kelman's lab (Institute for Bioscience and Biotechnology Research, University of Maryland).

2.3.2 Proteins expression and purification

The full-length MCM proteins harboring the Phe172 to Ile (F172I) and the Phe172 to Val (F172V) substitutions were generated by Dr. John Schimenti laboratory using QuickChange Site-Directed Mutagenesis Kits (Stratagene) using the wild-type full-length MCM in the pET-16b vector (Novagen) (23) as a template. All constructs contain a N-terminal His10-tag. All proteins were expressed in *Escherichia coli* BL21-DE3 cells at 24°C and induced with IPTG at a concentration of 1.0 mM. Proteins were expressed in *E. coli* cells BL21(DE3) at 37°C in the presence of appropriate antibiotics. When the culture reached an *A* 600 of 0.6, protein expression was induced by incubation in the presence of 1.0 mM IPTG overnight at 24°C after which time the cells were harvested. Cell lysates were prepared by sonication in lysis buffer containing 20mM Tris-HCl (pH 8.0), 500mM NaCl, 5mM Imidazole, 10% glycerol and protease inhibitor cocktail (containing Pepstatin A, Leupeptin, PMSF, and Benzamidine). After centrifugation for 30 min at 13000 rpm, extracts were incubated with Ni-NTA agarose column (Qiagen) for 2 h at 4 °C and washed with 10 column volumes of washing buffer (20mM Tris-HCl (pH 8.0), 500mM NaCl, 10% glycerol and protease inhibitor cocktail) containing 10 mM imidazole, and then 5 column volumes of Washing Buffer containing 100 mM imidazole. Then targeted proteins were eluted from the column with 5 column volumes of Elution Buffer (Washing Buffer containing 500 mM imidazole).

The wild-type or F172I mutant of the N-terminal part of mtMCM (residues 2–286) were cloned into pGEX-4T-1 plasmid (Amersham) between the BamHI and XhoI sites, using the plasmids containing the wild type or F172I mutant genes as templates for PCR. The proteins were overexpressed in *E. coli* BL21-DE3 cells at 24°C. The cells were lysed in buffer containing 50 mM Tris-HCl (pH 8.0), 250 mM NaCl and 10% glycerol and clarified by centrifugation. The lysate was incubated with Glutathione Sepharose 4 fast flow resin (GE Healthcare). Following extensive washing with same buffer, the proteins were cleaved from the resin using thrombin at 4°C in buffer containing 20 mM Tris-HCl (pH 8.5), 150 mM NaCl, and 2.5 mM CaCl2. The flow through was collected, concentrated and purified on Superdex-200 column (GE Healthcare).

All purified proteins used for the study were extensively dialyzed against buffer containing 20mM Tris-HCl (pH 8.0), 100mM NaCl, 0.5mM EDTA, 4mM DTT, and 10% glycerol. Protein concentrations were determined with Bradford (Bio-Rad) with bovine serum albumin (BSA) as the standard. Proteins were aliquoted, flash-frozen in liquid nitrogen and stored at -80°C.

2.3.3 Gel filtration analysis

The Gel filtration analysis of the full-length mtMCM wild-type and mutant proteins were performed at 25°C using Superose-6 gel filtration column (GE Healthcare) with buffer containing 50 mM Tris (pH 8.0), 250 mM NaCl, and 10% Glycerol. Prior to analysis the proteins were diluted to a concentration of 1mg/ml and incubated for 16 hr at 25°C. 150 µl samples were fractionated on the column.

The gel filtration of the N-terminal part of the wild-type and mutant mtMCM proteins were performed at 4°C using a Superdex-200 gel filtration column (GE Healthcare) with buffer containing 50 mM Tris (pH 8.0), 250 mM NaCl, and 10% Glycerol. Prior to analysis the proteins were diluted to a concentration of 2mg/ml and incubated for 16 hr at 4°C. 150 µl samples were fractionated on the column. (The gel filtration experiments of the N-terminal proteins were also performed in exactly the same conditions as full-length proteins, in which the results are consistent to the conditions listed above. The data presented in this thesis was under the conditions described above.)

2.3.4 Light scattering

The solution molecular masses of both proteins were determined using size exclusion chromatography with in-line multi angle light scattering. A 1200 series HPLC system (Agilent Technologies, Santa Clara, CA, USA) with a Shodex KW-802.5 column (Showa Denko K.K., Japan) was used to separate components of the solutions. The flow rate was 0.5 ml/min and the mobile phase was 0.01 M phosphate buffer (pH 6.5), 0.1 M ammonium sulfate, 0.1 mM EDTA and 0.01 % sodium azide. The light scattering was measured with a miniDawn Treos (Wyatt Technology, Santa Barbara, CA, USA); the protein concentration was measured with an Optilab rEX differential refractometer (Wyatt Technology, Santa Barbara, CA, USA). The proteins were diluted with the column buffer to the final concentration of 11 mg/ml (wild-type N-mtMCM) and 13 mg/ml (F172I N-mtMCM). 20 µl of each protein was used for the analysis. Molecular weights were calculated using Wyatt's

ASTRA software.

2.3.5 Circular dichroism

Circular dichroism (CD) measurements were performed with a Model J-720 Jasco Spectropolarimeter using quartz cells with a 0.005 cm path length at 24°C. The proteins used for CD were prepared as described above with concentrations of proteins ranging from 10-30 mM. Far-UV wavelength scans were recorded at room temperature from 260 to 200 nm. Averages for three CD spectra were presented. Ellipticity results were expressed as mean residue ellipticity[θ] = degrees x centimeter^2 x decimole^{-1}.

2.3.6 Helicase assay

a) Preparation of substrates for Helicase Assays

Substrates for helicase assays were generated as described below (following standard procedures referred in (24)) using the DF25 (5′-CCGACGTGCCAGGCCGACGCGTCCC-3′) and DF50 (5′-GGGACGCGTCGGCCTGGCACGTCGGCCGCTGCGGCCAGGCACC CGATGGC-3′) oligonucleotides to make the flat substrate and DF61 (5′-TTTGTTTGTTTGTTTGTTTGTTTGTTTGTTTGTTTGCCGACG TGCCAGGCCGACGCGTCCC-3′) and DF74 (5′-GGGACGCGTCGGCCTGGCACGTCGGCCGCTGCGGCCAGGCACC CGATGGCGTTTGTTTGTTTGTTTGTTTGTTT-3′) oligolucleotides to make the forked substrate.

The helicase substrates were made by labeling one DNA strand of each oligonucleotides at the 5' end using [γ-^{32}P]ATP with T4

polynucleotide kinase and adding EDTA of a final concentration of 25 mM to stop the labeling reactions. The labeled oligonucleotide was hybridized to the other oligonucleotide (usually at a 1:1 molar ratio) in a buffer containing 40 mM Hepes-NaOH (pH 7.5) and 50 mM NaCl by heating to 100 °C for 3 min followed by slow cooling to 25 °C. Unincorporated [γ-^{32}P]ATP and unannealed oligonucleotides were removed using the following procedure. After hybridization, a 6x loading buffer (0.1% xylene cyanol, 0.1% bromophenol blue, and 50% glycerol) was added to a final concentration of 1x, and the mixture was electrophoresed through an 8% native polyacrylamide gel for 1 h at 100 V in 0.5x TBE (45 mM Tris, 4.5 mM boric acid, 0.5 mM EDTA). The desired substrates were located by autoradiography, and the products were excised from the gel, which was sliced into small pieces and incubated at 37 °C for 2 h in 3 volumes of an elution buffer containing 0.5 M ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA (pH 8.0). After centrifugation, the supernatant was collected, and the insoluble material was extracted once more with elution buffer. Following centrifugation, both supernatant fractions were combined, and DNA substrates were then ethanol precipitated and dissolved in TE (10 mM Tris-HCl (pH 7.5), 1 mM EDTA).

b) DNA Helicase assay

DNA helicase activity assays were measured in reaction mixtures (15 ml) containing 20 mM Tris-HCl (pH 8.5), 2 mM DTT, 10 mM $MgCl_2$, 100 mg/ml bovine serum albumin (BSA), 3.33 mM ATP, 10 fmol (0.66 nM) (as monomer) of ^{32}P-labeled substrate and 10, 30 or 90 ng of full-length mtMCM wild-type or mutant proteins. Following incubation at 60°C for 1

hr the reactions were stopped by adding 5 ml of loading buffer containing 1% SDS, 100 mM EDTA, 0.1% xylene cyanol, 0.1% bromophenol blue and 50% glycerol. Aliquots (10 ml) were loaded onto an 8% native polyacrylamide gel in 0.5x TBE (44.5mM Tris, 44.5mM boric acid and 1 mM EDTA) and electrophoresed for 40 min at 180V. Gels were visualized and quantitated by phosphorimaging. All helicase experiments were repeated three times, and their averages with SD are shown in the figure.

The doping experiment was performed as described above with the flat substrate using 90 ng of wild-type MCM and mutant proteins in ratios of 1:5, 3:3, 5:1 and 6:0. Experiments were repeated three times, and their averages with SD are shown in the figure.

2.3.7 Statistical Model for Doping Assay

Addition of mutant protein to reaction allows the formation of hetero-oligomers between the WT and mutant enzymes. The relative abundance of each species should follow a binomial distribution if non-cooperative binding properties are obtained between two forms of mtMCMs. So Equation (1) was used to calculate the possibility of each hexameric species present in the reaction mixture at any given ratio of wild-type and mutant protein, where "p" is the proportion of the amount of wild-type protein out of total amount of proteins; "x" is the number of wild-type protein molecule in a hexamer, which can take value of 0,1,2,3,4,5,6; then the equation describing the distribution of X is,

$$\Pr(X = x) = \binom{6}{x} p^x (1-p)^{6-x} \qquad (1)$$

This equation is only true when $0<p<1$. If there is only wild-type protein in the reaction, then Pr(X=6)=1, and Pr(X=0)=Pr(X=1)=Pr(X=2)=Pr(X=3) =Pr(X=4)=Pr(X=5)=0; If there is no wild-type protein in the reaction (mutant protein only), then Pr(X=0)=1, and Pr(X=1)=Pr(X=2)=Pr(X=3)= Pr(X=4)=Pr(X=5)= Pr(X=6)=0.

2.3.8 ATPase assay

ATPase activity was measured in reaction mixtures (15 µl) containing 25 mM Hepes-NaOH (pH 7.5), 1 mM dithiothreitol, 5 mM $MgCl_2$, 100 µg/ml BSA, and 1500 pmol of [γ-^{32}P]ATP (3000 Ci/mmol; GE healthcare) and 30 or 90 ng full-length mtMCM wild-type and mutant proteins in the absence or presence of 50 ng ssDNA (5'-GGCAGATAACAGTTGTCCTGGAGAACGACCTGGTTGACACCCTC ACACCC - 3'). After incubation at 60°C for 60 min, samples were placed on ice, then an aliquot (1 µl) was spotted onto a polyethyleneimine cellulose thin layer plate, and ATP and Pi were separated by chromatography in 1 M formic acid and 0.5 M lithium chloride. The extent of ATP hydrolysis was quantitated by phosphorimager analysis. All ATPase assays were repeated three times, and their averages with SD are shown in the figure.

2.3.9 DNA binding assay

Fluorescence anisotropy measurements were performed at 25°C using a Varian Cary Eclipse Spectrofluorometer equipped with an autopolarizer in a L-format, using a 10 mm path length cuvette with a starting volume of 1 ml. A 50-mer ssDNA (5'-

CGCAGATAACAGTTGTCCTGGAGAACGACCTGGTTGACACCCTC ACACCC-3') labeled at the 5'-labeled with ROX was used for the study. The initial reaction mixture contained 25 mM Hepes-NaOH (pH 7.5), 2 mM DTT, 5 mM $MgCl_2$ and 30 nM DNA. After each addition of proteins, the anisotropy value was monitored for 10-15 minutes until it reach the equilibrium. Each reading was measured with a setting of 10 sec integration. An average of several readings in equilibrium was recorded as the anisotropy value at each protein concentration. The DNA was excited at 588 nm and emission spectrum was set at 608 nm. The G-factor was acquired once at the beginning of each experiment.

The binding constant (K_d) was determined using GraFit version 5.0.1 (Erithacus software), using the following equation for fluorescent polarization anisotropy experiments [17,18]; where A is the change in anisotropy, A_T is the total anisotropy change, E is the enzyme concentration at each titration point, D_T is the total concentration of DNA (assuming it is constant at 30 nM), Kd is the dissociation constant for the binding isotherm and nH is the fitted Hill coefficient. The experiments were repeated twice and their averages with SD are shown in the figure.

$$\Delta A = \Delta A_T \left[\frac{\left(E/K_d\right)^{nH}}{1+\left(E/K_d\right)^{nH}} \right] \qquad (2)$$

2.3.10 Energy minimization

A Hexameric wild-type N-terminal model was constructed from the crystal structure of wild-type mtMCM (Ref. 15, PDB code: 1LTL).

Hexameric F172I and F172V mutants were built by changing Phe172 to Ile172 and Val172, respectively. As the side chains of Ile and Val residues are shorter than that of Phe, they were visually modeled to conformations with least close contacts to surrounding atoms and with maximum overlaps over the phenyl ring of Phe. To further remove possible close contacts introduced by the manual manipulation, conjugate-gradient energy minimization implemented in the program CNS (version 1.2) (37,38) was performed on the hexameric wild-type as well as mutant models. Calculations converged after 200 steps of energy minimization. To determine the structural impact by the F172I/F172V mutations on the hexameric assembly of mtMCM, residue 172 and four closest residues (Ser103, Ile106, Thr222, and Phe231) were extracted from its neighboring monomer for detailed energy analysis. Energy for each of the five-residue models described above was calculated by using CNS and was compared with each other as shown in the figure.

2.3.11 Processivity assay

Three forked substrates with increasing duplex lengths (Suppl. Table 1A,B,C) were generated and used for helicase assay as previously described, to investigate the processivity ability of full-length wild-type and mutant helicases. 6% native polyacrylamide gels were used for electrophoresis to display wound and unwound substrates. Experiments were repeated three times, and their averages with SD are shown in the figure.

2.3.12 TrpY protein replacement assay

This assay was performed by Nozomi Sakakibara. A helicase substrate containing the regulatory region, including the four TRP-box sequences, from upstream of the M. thermautotrophicus trpEGCFBAD operon (Suppl. Table 1D) was used for unwinding assay. The ability of full-length wild-type and mutant helicases to unwind this DNA substrate were assayed when protein-free or bound by TrpY. Each protein was assayed at two different concentrations, 50ng and 150ng per reaction. TrpY binding substrate was prepared and helicase assays were done as previously described (36). Experiments were repeated three times, and their averages with SD are shown in the figure.

2.4 Results

2.4.1 Phe172 is needed for efficient helicase activity.

The MCM complex is thought to function as the replicative helicase in archaea and eukarya. Thus, although the mice harboring the Chaos3 mutation are viable, they show an increase in chromosomal instability suggesting a defect in DNA replication (5,6). To determine how the Chaos3 mutation affects MCM helicase activity the conserved Phe residue (Phe172) was substituted with either Ile or Val in the mtMCM protein (Fig. 2.2). When a forked DNA substrate containing both 3' and 5' overhang ssDNA was used, the mutants and wild-type enzymes exhibit similar helicase activity (Fig. 2.2A compare lanes 6-11 to lanes 3-5, see also panel B). As mice harboring the MCM^{Chaos3} mutation are viable, one would expect the enzyme to retain some activity (all six Mcm2-7 proteins are essential for viability). It was previously shown that the archaeal MCM

proteins, including mtMCM, were more active on forked substrate compared to substrate that contains only a 3'-overhang ssDNA region (flat substrate) (27). Thus, a flat substrate was also used to determine the effect of the mutations as it may indicate impaired activity. As shown in Figure 2.2 C and D, while the substitution of Val for Phe does not have a major effect on helicase activity (Fig. 2.2C compare lanes 9-11 to lanes 3-5, see also panel D), the Phe to Ile substitution resulted in very limited helicase activity compared to the wild-type enzyme (Fig. 2.2 compare lanes 6-8 to lanes 9-11, see also panel D). As a control, a mutant protein harboring a mutation of a conserved Lys needed for ATP binding and hydrolysis (K325A) was used. This mutant was previously shown to be devoid of helicase activity (28,29) as is also found here (Fig. 2.2C lane 12). These results suggest that the in vivo observation made with the Chaos3 mutation is due to impaired helicase activity by the helicase.

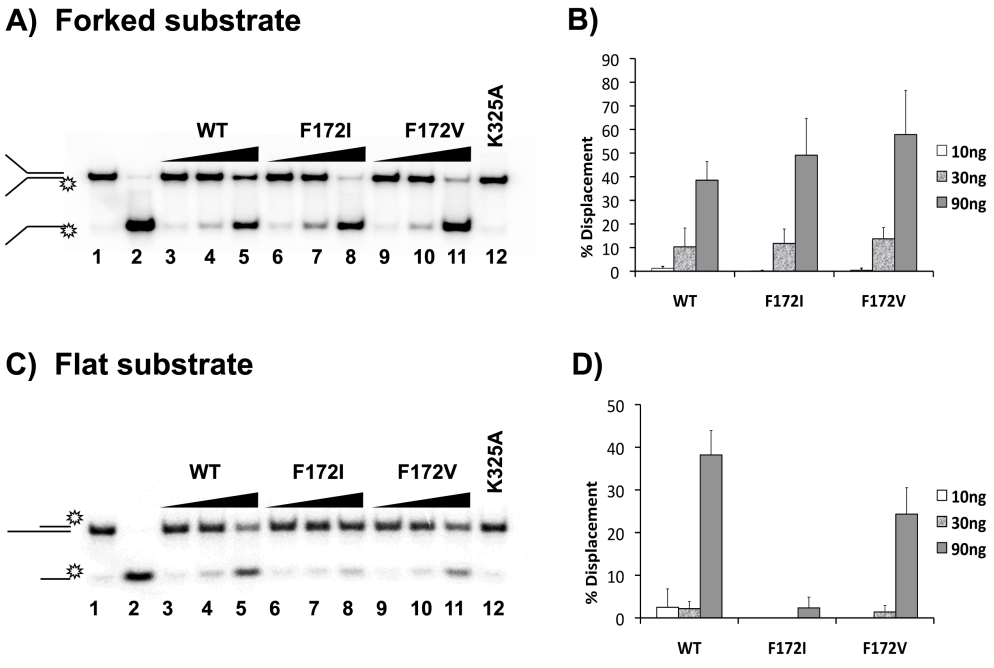


Figure 2.2 mtMCM with Ile substitution for Phe at position 172 exhibits impaired helicase activity. Helicase assays were performed as described in "Materials and

Methods" using 10 fmol of flat (A and B) or forked (C and D) substrate. Representative gels are shown in panels A and C and the summary of three independent experiments with standard deviation are shown in panels B and D. In panels A and C lane 1, substrate only; lane 2, boiled substrate; lanes 3, 6 and 9, 10 ng of proteins; lanes 4, 7 and 10, 30 ng of proteins; lanes 5, 8, 11 and 12, 90 ng of proteins. Lanes 3-5, wild-type mtMCM protein; lanes 6-8, F172I mutant mtMCM protein; lanes 9-11, F172V mutant mtMCM protein; lane 12, K325A mutant mtMCM protein.

Another approach to demonstrate that the F172I mutation impairs helicase activity is by performing a doping experiment. In this assay, the ratio between the wild-type and mutant proteins decreases in each reaction, resulting in heterocomplexes with different proportions of wild-type and mutant protein. If a mutant subunit affects the overall helicase activity then as the ratio between the mutant and wild-type protein increases in the assay, the enzyme will become less active. The addition of F172V to wild type protein did not have a major effect on the helicase activity (Fig. 2.3), as expected from the helicase activity of the mutant protein alone (Fig. 2.2). In contrast, the formation of heterocomplexes containing subunits of the F172I mutant and wild type proteins resulted in a substantial reduction in helicase activity (Fig. 2.3). Consistent with the helicase assay results of Figure 2.2, the mutant enzyme retained ~5% of helicase activity even when only the mutant protein was used. However, reduction in activity was evident when the ratio between the wild type and mutant protein (F172I) was 5:1 suggesting that a single mutant protein in the hexameric ring may be sufficient to affect helicase activity.

If one assumes free exchange between hexamers, the number of hexamers that contain only wild-type proteins should follow a binomial distribution. Therefore, when the ratio between wild-type and mutant is 5:1, 33.5 percent of hexamers should contain only wild-type proteins and 1.5 percent when the ratio is 3:3. Given the maximum wild-type unwinding data of 55% and assuming that a single Walker-A mutant

protein complex with 5 wild-type enzymes will abolish helicase activity, one would expect 18.5% and 1% unwinding when the ratios between wild-type and mutant protein (K325A) are 5:1 and 3:3, respectively. This is not the case, however, higher activity was observed (Fig. 2.3). If both the wild-type and mutant proteins were added to the reaction as stable hexamers, then it is possible that an hour of incubation was not sufficient to reach equilibrium. It is also possible that one subunit of the hexamer containing a Walker-A mutant is not sufficient to abolish helicase activity, as was suggested for S. solfataricus (30). In comparison with the observations made with the K325A mutant, the reduction in helicase activity by the F172I mutant protein is less dramatic, suggesting F172I is a milder mutation than K325A.

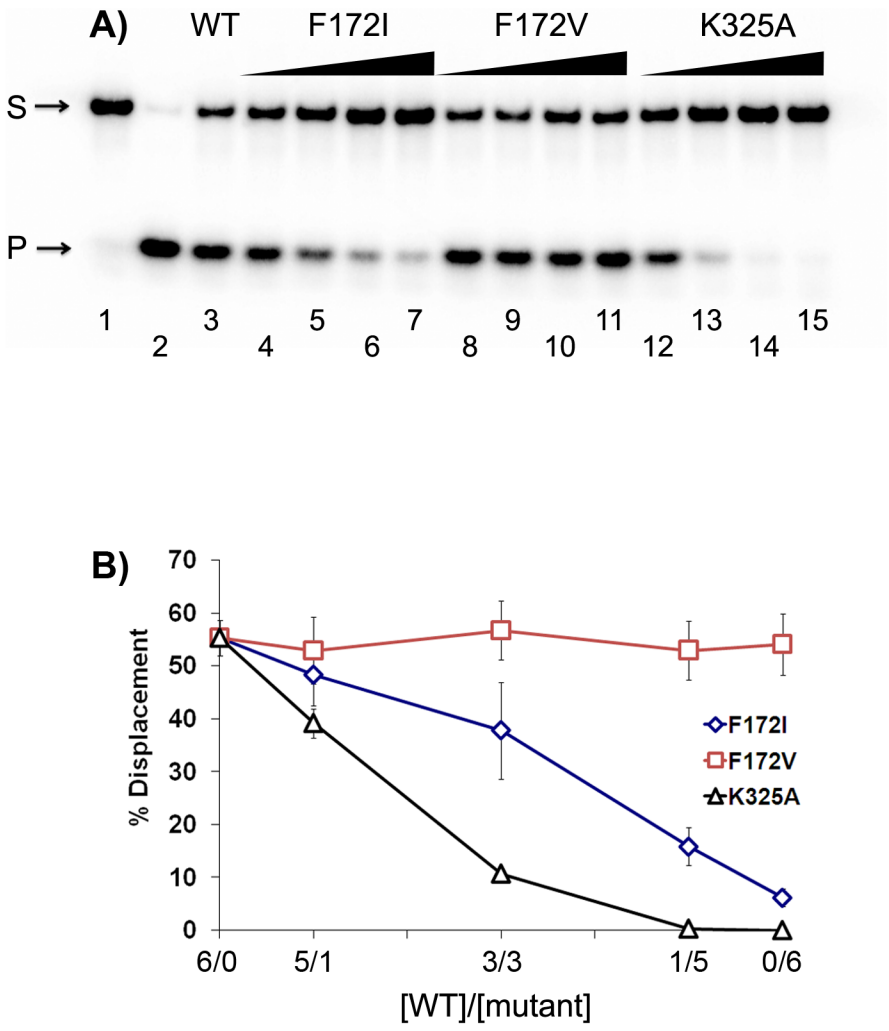


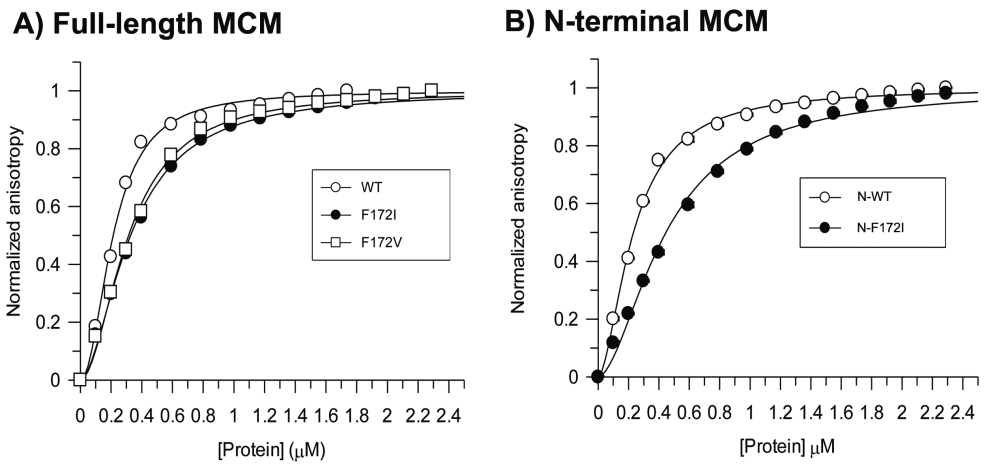
Figure 2.3 mtMCM F172I mutant protein inhibits helicase activity when complexed with the wild-type enzyme. Helicase assays were performed as described in "Materials and Methods" using 10 fmol of flat substrate and 90 ng of protein. A representative gel is shown in panel A and the result of three independent experiments with standard deviation is shown in panel B. In panel A lane 1, substrate only; lane 2, boiled substrate; lane 3, wild-type only; lanes 4, 8 and 12, the ratio between the wild-type and mutant proteins is 5:1; lanes 5, 9 and 13, ratio of 3:3; lanes 6, 10 and 14, ratio of 1:5; lanes 7, 11 and 15, mutant protein only. Lanes 4-7, F172I; lanes 8-11, F172V; lanes 12-15, K325A.

In eukarya the active helicase is a heterohexamer in which only one copy of MCM4 is present. Therefore there is only one MCM4^{Chaos3}

mutant protein in each helicase. The data presented in Figure 2.3 suggest that the presence of even one mutant protein in the hexameric ring is sufficient to reduce helicase activity. This observation may explain why a mutation in a single MCM in eukarya is sufficient to affect helicase activity.

2.4.2 Phe172 is required for efficient DNA binding by MCM

An active helicase binds DNA and translocates along it in an ATP-dependent manner. The experiments described above (Fig. 2.2 and 2.3) demonstrate that the F172I mutant protein has impaired helicase activity. The impaired activity can be a result of reduced affinity to DNA and/or reduced ATPase activity by the helicase. Therefore, to get a better understanding of the Chaos3 mutation, the mutant proteins (F172I and F172V) were evaluated for their ability to bind DNA and to hydrolyze ATP.



C) Dissociation constants (nM)

Protein	Kd
FL-mtMCM (wild-type)	42.0±6.5
FL-mtMCM (F172I)	137.7±9.2
FL-mtMCM (F172V)	111.1±8.7
N-mtMCM (wild-type)	81.8±5.9
N-mtMCM (F172I)	242.8±14.7

Figure 2.4 Mutations of Phe 172 do not significantly affect DNA binding by mtMCM. The K_d values of the interactions between mtMCM wild-type and mutant proteins and DNA were measured using Fluorescence polarization anisotropy (FPA) analysis as described under "Materials and Methods" in the presence of 30 nM Rox-labeled ssDNA. The anisotropy change was measured as the proteins were titrated into the reaction mixture. Panel A, full-length enzyme; panel B, N-terminal part of the molecule. Data were analyzed as described in "Materials and Methods" and the K_d values are shown in panel C.

Although mtMCM interact with DNA mainly via the N-terminal part of the molecule (15,17,18) it was shown that the catalytic domain also participates in DNA binding (31). Thus the ability of the mutation in the full-length protein (Fig. 2.4A) or the N-terminal part (Fig. 2.4B) to bind DNA was determined. As expected, the wild-type full-length enzyme binds DNA with higher affinity (Kd of 42nM) in comparison to the N-

terminal part alone (Kd of 82nM) (Fig. 2.4C). When the mutations are presence in the full-length background both enzymes (F172I and F172V) bind ssDNA with slightly lower affinities (Kd of 110 and 140nM, respectively) in comparison to the wild-type enzyme (Fig. 2.4A and 2.C). Similarly, the F172I mutation on the background of the N-terminal part of the molecule also resulted in reduced affinity to ssDNA (Kd of 243nM) (Fig. 2.4B and 2.C). However, the reduced affinity is rather small and similar in both mutant (Kd of 138 and 111 for F172I and F172V, respectively) and thus this reduction is unlikely to affect helicase activity observed with F172I (Fig. 2.2 and 2.3). The minimal effect of mutation on DNA binding is not unexpected as Phe172 is not located near the motifs involved in DNA binding (the zinc-finger and β-hairpin motifs) (Fig. 2.1A) (15,17,18).

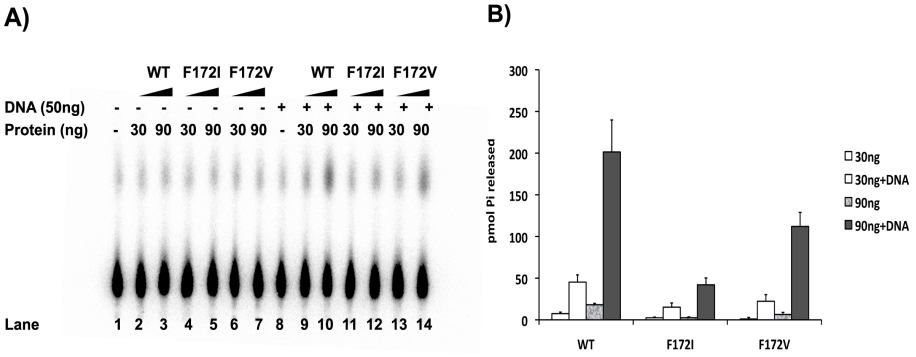


Figure 2.5 Mutations of Phe 172 reduce ATPase activity by mtMCM. ATPase activity of wild-type and mutant mtMCM proteins in the absence and presence of ssDNA was determined as described in "Materials and Methods" using 30 and 90 ng of proteins (as monomer) in the absence (lanes 1-7) or presence (lanes 4-14) of 50ng of 50 mer ssDNA. A representative gel is shown in panel A; the summary of three independent experiments with standard deviation is shown in panel B.

Next, the ability of the mutant protein to hydrolyze ATP was determined. Although the mutations are at the N-terminal part of the

molecule and not in the AAA+ catalytic domains it was shown that mutations at the NCL of the N-terminal part of MCM affect the ATPase activity of the catalytic region (20,21). As shown in Figure 2.5 both mutant proteins (F172I and F172V) exhibit some reduction in basal ATPase activity compared to the wild-type MCM protein (Fig. 2.5B). However, DNA stimulates the ATPase activity of both mutant proteins to the same extent as the stimulation of the wild-type enzyme (Fig. 2.5). The reduced ATPase activity of the F172I mutant protein may account for the reduced helicase activity. The reduced ATPase activity, however, cannot explain the difference in helicase activity observed on forked and flat substrates (Fig. 2.2). Thus, it is unlikely that the effect of the Chaos3 mutation on helicase activity is via reduction in DNA binding and/or ATPase activity by the mutant enzymes.

2.4.3 Phe172 participates in MCM hexamerization.

In addition to affecting DNA binding and ATPase activities, mutation of Phe 172 may also affect the structure and oligomeric state of the protein which in turn may affect the enzymatic activity. It is well established that the active helicase has a hexameric form ((32) and references therein). Thus, it is possible that the F172I mutation affects the structure or stability of the hexameric ring. Phe172 from one monomer is in close proximity to the neighboring subunit (Fig. 2.1C) supporting the possibility of the mutation affecting hexamer integrity. To evaluate this possibility, gel-filtration analyses of the full-length F172I and F172V proteins were performed (Fig. 2.6). It was found that the full-length versions of all three proteins form similar oligomeric structures (Fig. 2.6A). However, in order to detect the protein on gel filtration, high protein concentrations (13 mM) were used. It was shown that protein concentration plays a major role in

MCM oligomeric state (32). One way to determine whether the structure of the mutant proteins is compromised even if high concentrations are used is to analyze only the N-terminal part of the molecule. Biochemical and structural studies showed that the N-terminal part of the wild-type enzyme is hexameric (Fig. 2.6B) (15,22).

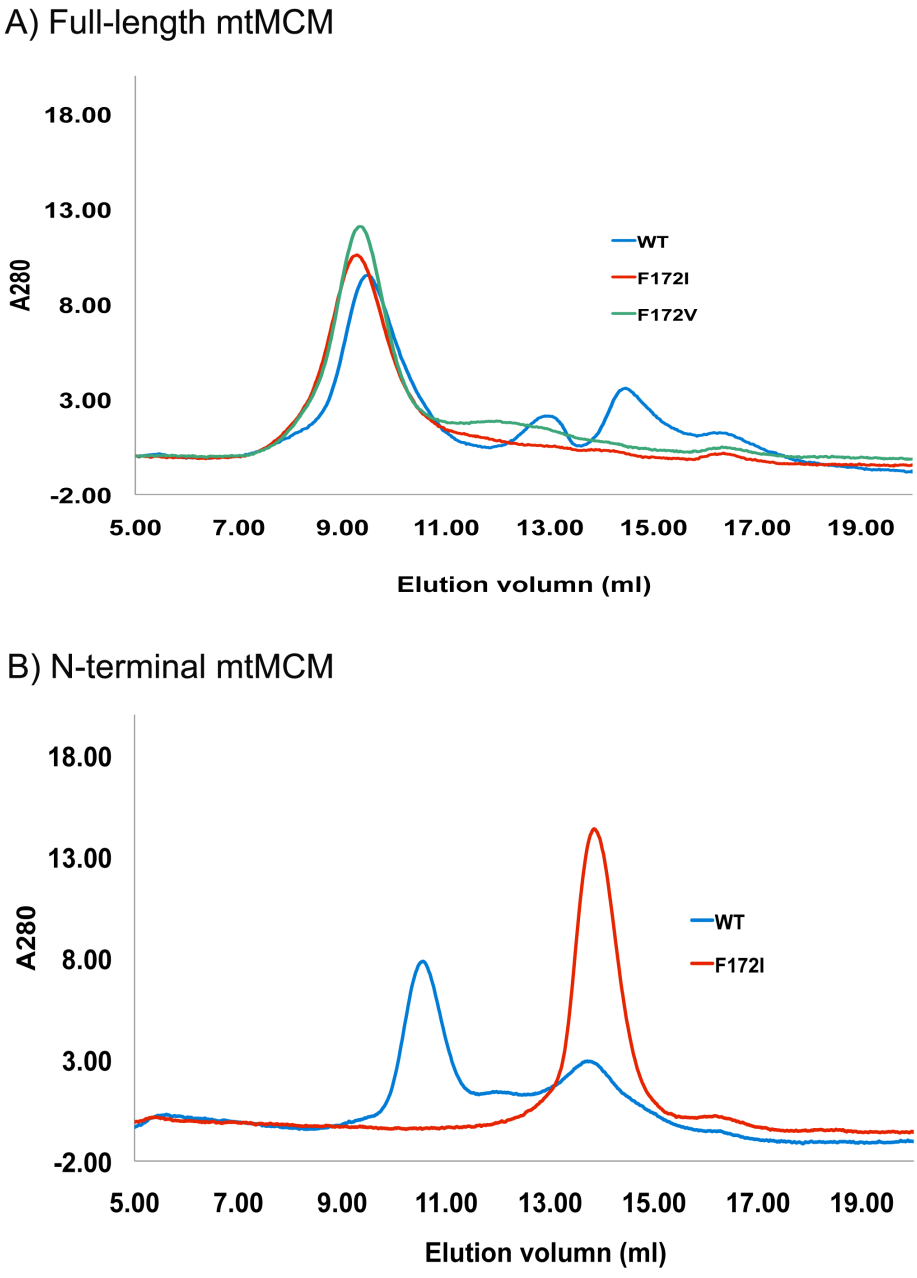


Figure 2.6 Mutations of Phe 172 affect the oligomeric structure of mtMCM. Gel-filtration analysis of the full-length (A) or N-terminal part (B) of wild-type and mutant MCM proteins. One hundred micrograms of protein were loaded onto a superose-6 (A) or superdex-200 (B) gel filtration column and analyzed as described in "Materials and Methods".

The F172I mutant, on the other hand, is monomeric (Fig. 2.6B) although the overall secondary structure was not changed (Fig. 2.7). These results suggest that MCM proteins harboring the mutation form less stable hexamers in comparison to the wild-type enzyme.

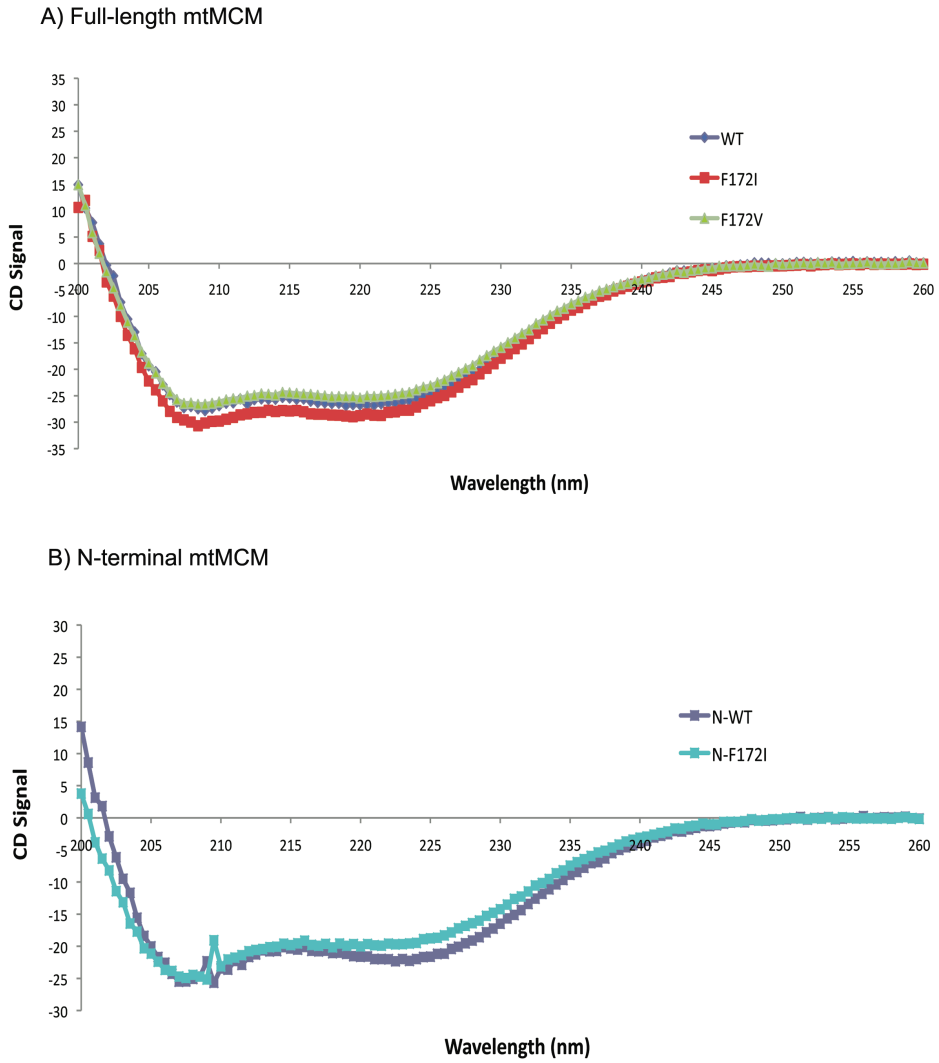


Figure 2.7 Mutations of Phe172 does not affect the over all secondary structure of mtMCM. Circular dichroism (CD) spectra of wild-type and mutant mtMCM proteins. Each spectra was normalized to mean residue ellipticity [θ] = degrees x centimeter^2 x decimole^{-1} according to the concentrations of protein determined as described in "Materials and Methods". The spectra were taken at 24°C.

Taken together, the results suggest that the substitution of Ile for Phe at position 172 affects the stability of the MCM hexamer and may contribute to its helicase activity.

2.4.4 Structure simulation suggests a thermodynamically less stable model for F172I mutant protein

Another approach to evaluate the stability of the enzyme is by using energy minimization analysis, which can indicate the impact of the mutation on the structure. Three N-terminal mtMCM protein models (wild-type, F172I and F172V) were built as described in "Materials and Methods" (Fig. 2.8). Amino acid 172 and its four closest residues (Ser103, Ile106, Thr222, and Phe231) from the neighboring subunit were chosen for detailed energy analysis. By comparing the total energy of these three 5-residue models, it was found that the model with F172I mutation has higher energy (-61.942 cal/mol) than the wild type protein (-80.885 cal/mol), suggesting that it is thermodynamically less stable. The F172V model appears to have slightly lower energy (-93.589 cal/mol) than the wild type protein, which might explain their similar activity levels for all enzymatic reactions studied. Taken together, the results suggest that the substitution of Ile for Phe at position 172 affects the stability of the MCM hexamer and thus may contribute to the reduction of its helicase activity.

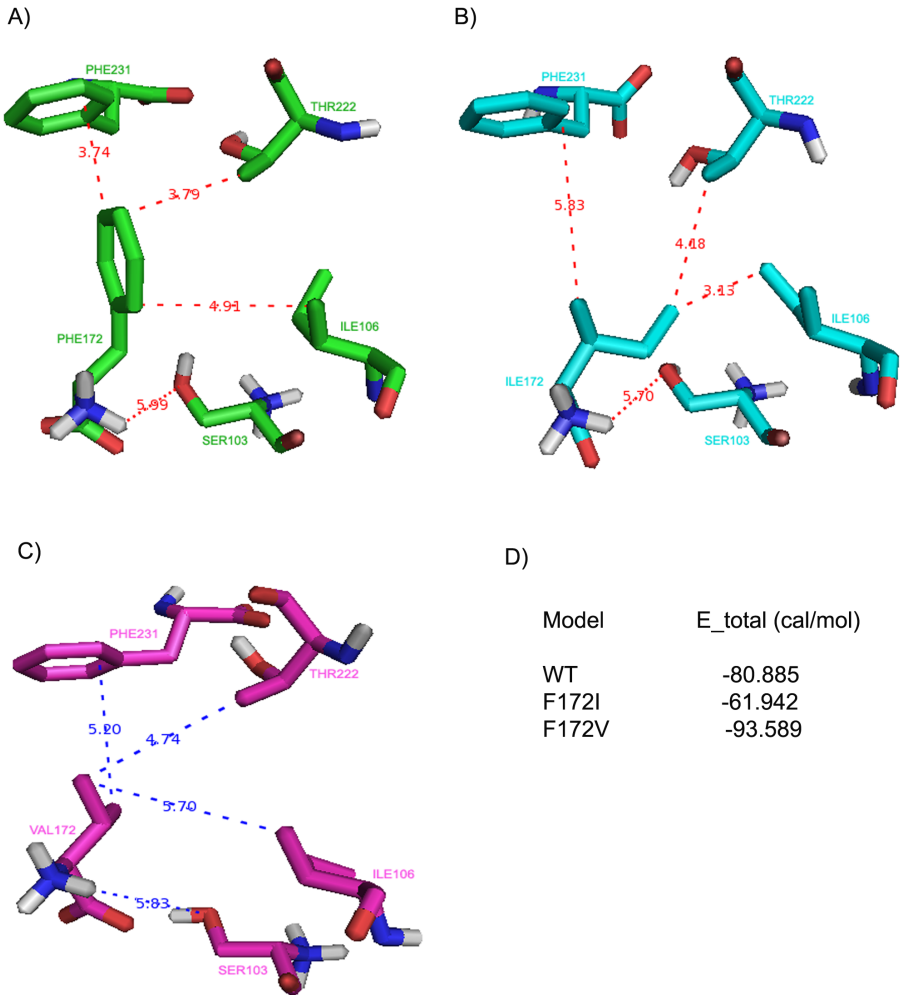


Figure 2.8 5-residue simulation model of N-terminal wild-type and mutant proteins. A) Wild-type 5-residue model. The distance between Phe172 and other 4 residues are shown with red dot lines. B) F172I mutant 5-residue model. The distance between Ile172 and other 4 residues are shown with red dot lines. C) F172V 5-residue model, the distance between Val172 and other 4 residues are shown with blue dot lines. The units for the distances are all in Å. Total energy of each protein model was calculated based on the above coordinates. D) Total energy comparisons of three 5-residue simulation models. The program CNS was used with one-step energy minimization to calculate the total energy of each model system.

2.4.5 F172I mutation does not affect processivity activity

The ability of both wild-type and mutant mtMCM proteins to process forked substrates with increased duplex regions was assayed. Results show that mutations of Phe172 do not apparently affect helicase processivity activity. All three proteins can unwind forked DNA substrates up to 80bp duplex length. But the helicase activity of F172I mutant protein does get weaker when the duplex region of forked substrate increases, which is consistent with the results on helicase assay with flat substrate. The effect of F172I mutation on forked substrate with 26bp duplex (Suppl. Table 1A) is hard to see (Fig. 2.9A compare lanes 6-11 to lanes 3-5, see also panel D), as it is discussed in helicase assay previously. But the helicase activity of F172I mutant was apparently reduced when 55bp duplex substrate (Suppl. Table 1B) was used (Fig. 2.9B compare lanes 6-11 to lanes 3-5, see also panel D), and the effect was even greater on 80bp duplex substrate (Suppl. Table 1C) (Fig. 2.9C compare lanes 6-11 to lanes 3-5, see also panel D). Although this assay does not provide enough information to tell a difference on processivity ability of all three proteins, it does confirm that the F172I mutation does reduce the enzyme's helicase activity.

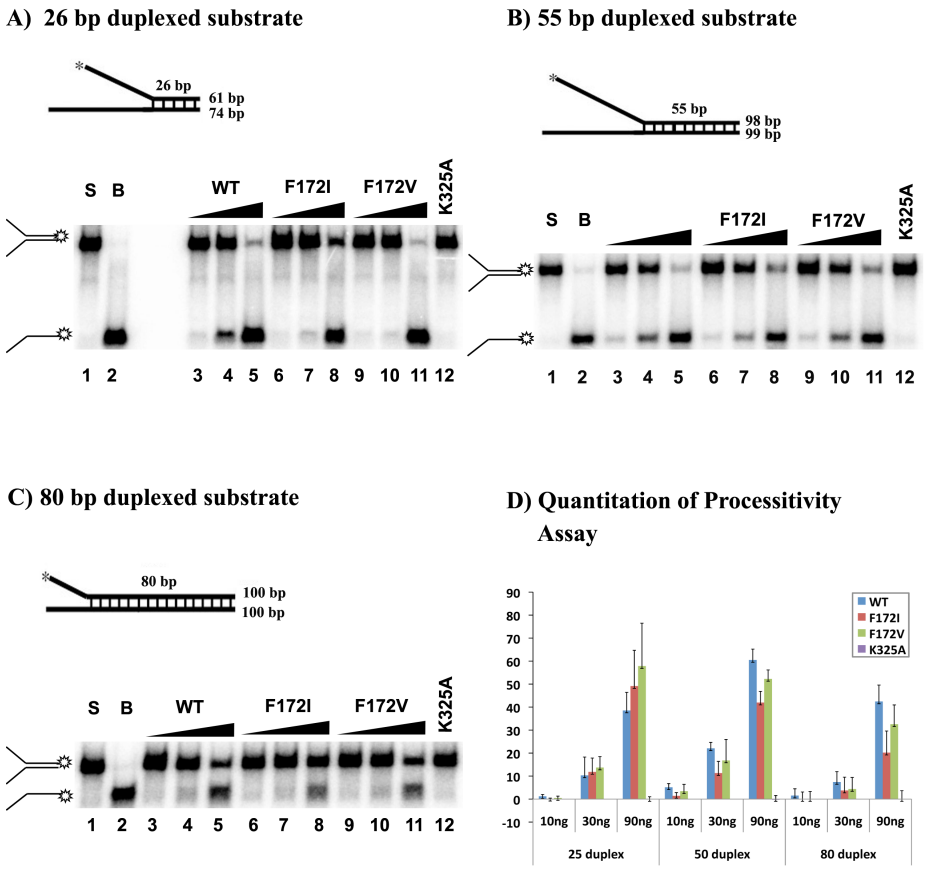


Figure 2.9 The processivity of mtMCM mutant helicase were not significantly affected. Mutant proteins can process forked substrate with up to 80 bp duplex, but the F172I mutant protein exhibits impaired helicase activity on long duplex forked substrates. Helicase assays were performed as described in "Materials and Methods" using three forked substrates with increasing duplex length. The substrate used for each assay was illustrated at the top of the figure with the location of the ^{32}P-label identified by the asterisk. Representative gels of assays using forked substrate with 26bp, 55bp and 80bp duplex length are shown respectively in panels A, B and C and the summary of three independent experiments with standard deviation are shown in panels D. In panels A, B and C lane 1, substrate only; lane 2, boiled substrate; lanes 3, 6 and 9, 10 ng of proteins; lanes 4, 7 and 10, 30 ng of proteins; lanes 5, 8, 11 and 12, 90 ng of proteins. Lanes 3-5, wild-type mtMCM protein; lanes 6-8, F172I mutant mtMCM protein; lanes 9-11, F172V mutant mtMCM protein; lane 12, K325A mutant mtMCM protein.

2.4.6 F172I mutation does not affect protein displacement activity

TrpY is a M. thermautotrophicus transcription repressor that binds specifically to TRP-box sequences (consensus TGTACA) that are located

upstream of the trpY, trpEGCFBAD, and trpB2 genes in the Methanothermobacter thermautotrophicus genome. The helicase substrate that contains 4 TrpY protein binding sites (Suppl. Table 1D) was therefore used to investigate the ability of mutant protein to replace TrpY and unwind the substrate (36).

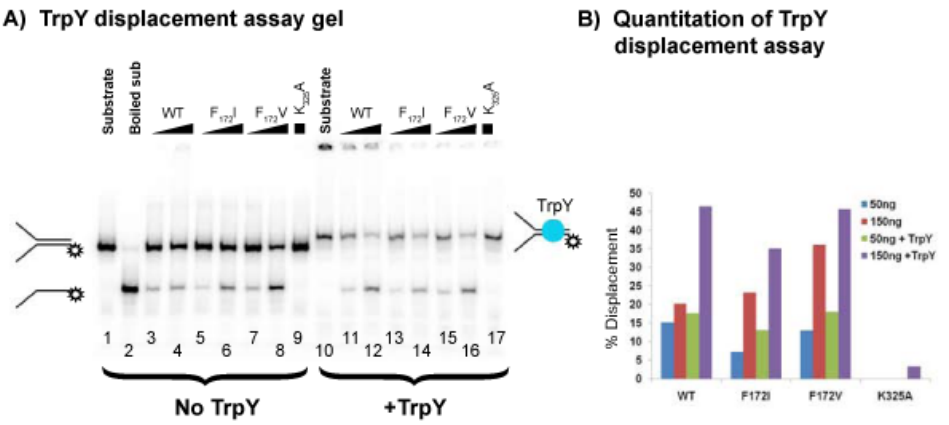


Figure 2.10 Mutations of Phe 172 do not affect protein displacement activity by mtMCM. The ability of the mutant mtMCMs to unwind DNA substrate was assayed as described in "Materials and Methods" using 50 ng and 150 ng of proteins (as monomer) with substrate protein-free (lanes 1-9) or TrpY bound (lanes 10-17). A representative gel is shown in panel A; the summary of three independent experiments with standard deviation is shown in panel B. In panel A, lanes 1 and 10, substrate only; lane 2, boiled substrate; lanes 3, 5, 7, 11, 13 and 15, 50 ng of proteins; lanes 4, 6, 8, 12, 14 and 16, 150 ng of proteins. Lanes 3-4 and 11-12, wild-type mtMCM protein; lanes 5-6 and 13-14, F172I mutant mtMCM protein; lanes 7-8 and 15-16, F172V mutant mtMCM protein; lane 9 and 17, 150 ng of K325A mutant mtMCM protein.

Results show that protein displacement activity of either F172I or F172V mutant was not abolished. But as expected to be weaker helicases, the unwinding efficiency of F172I mutants were reduced with either substrate (Fig. 2.10A compare lanes 5-8 to lanes 3-4 for protein free substrate and lanes13-16 to lanes 11-12 for TrpY bound substrate, see also panel B). This result is consistent with the above experiments. Again, as a negative control, neither the TrpY-bound substrate nor the protein-free

substrate was unwound by the MCM (K324A) mutant that lacked helicase activity (Fig. 2.10, lane 17 and 9).

2.5 Discussion

In this study, the MCM protein from the archaeon M. thermautotrophicus was used as a model to study the mouse Chaos3 mutation. The mutation in the mouse MCM4 protein is a substitution of a conserved Phe residue with Ile. Examination of the structure of the N-terminal part of mtMCM revealed that Phe172 is a part of a β strand (β7) located in domain C where it connects to domain B (Fig. 2.1) (15). In the hexameric structure, Phe172 is located at the interface between neighboring monomers (Fig. 2.1C) (15). The proximity of Phe172 from one monomer to another monomer in the hexameric structure suggests that the residue may play a role in hexamer formation. In the wild-type MCM protein, the phenyl ring of Phe172 interacts with the phenyl ring of Phe231 of the neighboring monomer (Fig. 2.11) by an edge-face π overlap. This interaction is lost when Phe172 is mutated to Ile or Val. These results are also consistent with the previously reported observation that domain C, where Phe172 is located, is needed for mtMCM hexamerization (22).

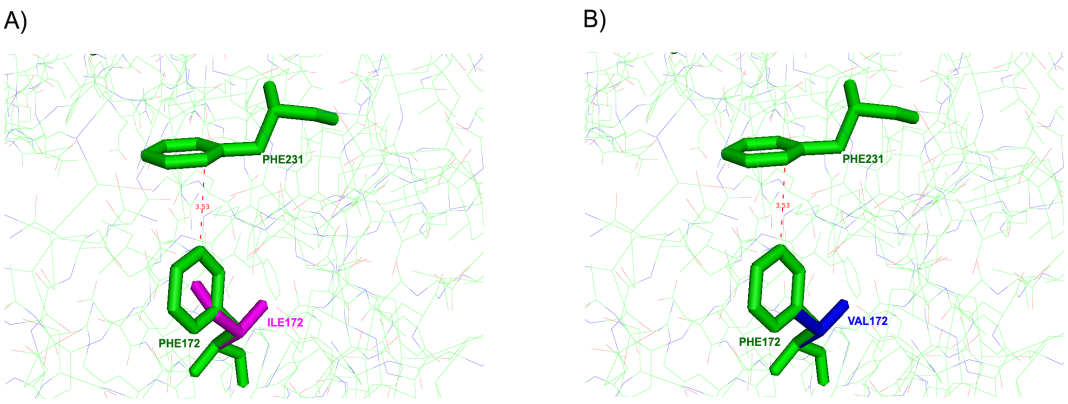


Figure 2.11 The role of residue 172 in hexamer formation. In the wild-type MCM protein, the

phenyl ring of Phe172 interacts with the phenyl ring of Phe231 of the neighboring monomer by an edge-face π overlap. This interaction is lost when Phe172 is mutated to Ile (A) or Val (B).

The effect of the mutation on hexamer stability can also explain the observation that while little effect on mtMCM helicase activity was observed when analyzed on a fork-like structure with both 3' and 5' overhang regions, very limited activity was detected on substrate with only a 3' overhang. It was previously shown that when provided with a fork-like substrate the archaeal MCM complex encircling the 3' ssDNA strand also interacted with the 5' overhang ssDNA region (33). Using the eukaryotic MCM complex it was shown that the processivity of the helicase is substantially increased when provided with a forked DNA substrate and it was shown that the helicase interacts with both single stranded strands (34). Thus, it is likely that although the hexamers harboring the F172I mutations are less stable the additional interactions with both ssDNA at the fork-like substrate enhances the number of contacts between the helicase and DNA thus enables the enzyme to unwind the duplex. This additional stabilization is missing in the flat substrate.

It was observed that in mouse cells harboring the Chaos3 mutation the stability of MCM4 and other MCM proteins was reduced (5). This may also result from reduced stability of the mutant MCM molecule. Studies have shown that in vivo most of the MCM proteins are part of complexes involving either all six proteins or several subsets of the proteins (for example see: (35)). Thus, it is possible that these complexes are protected from degradation. When the mutant MCM4 is expressed in vivo the complexes formed are less stable and thus may be more prone to degradation.

The data presented here suggest that the MCM complexes containing the $MCM4^{Chaos3}$ protein are less stable, providing a possible explanation for the observations made with mice harboring the mutation. It remains to be determined, however, why the most severe effect of the mutation is observed in females and how the lowered helicase activity and reduced stability of the

complex observed with the mutant enzyme results in chromosomal instability.

REFERENCE

1. Tye, B. K. (1999) *Annu Rev Biochem* **68**, 649-686
2. Tye, B. K., and Sawyer, S. L. (2000) *J Biol Chem* **275**, 34833-34836
3. Forsburg, S. L. (2004) *Microbiol Mol Biol Rev* **68**, 109-131
4. Bochman, M. L., and Schwacha, A. (2009) *Microbiol Mol Biol Rev* **73**, 652-683
5. Shima, N., Alcaraz, A., Liachko, I., Buske, T. R., Andrews, C. A., Munroe, R. J., Hartford, S. A., Tye, B. K., and Schimenti, J. C. (2007) *Nat Genet* **39**, 93-98
6. Shima, N., Buske, T. R., and Schimenti, J. C. (2007) *Cell Cycle* **6**, 1135-1140
7. Steere, N. A., Yamaguchi, S., Andrews, C. A., Liachko, I., Nakamura, T., and Shima, N. (2009) *Mutat Res* **666**, 74-78
8. Li, X. C., Schimenti, J. C., and Tye, B. K. (2009) *PLoS Biol* **7**, e1000161
9. Kelman, L. M., and Kelman, Z. (2003) *Mol Microbiol* **48**, 605-615
10. Sakakibara, N., Kelman, L. M., and Kelman, Z. (2009) *Mol Microbiol* **72**, 286-296
11. Duggin, I. G., and Bell, S. D. (2006) *J Biol Chem* **281**, 15029-15032
12. Costa, A., and Onesti, S. (2009) *Crit Rev Biochem Mol Biol* **44**, 326-342
13. Brewster, A. S., Wang, G., Yu, X., Greenleaf, W. B., Carazo, J. M., Tjajadi, M., Klein, M. G., and Chen, S. X. (2008) *Proc Natl Acad Sci U S A* **105**, 20191-20196
14. Bae, B., Chen, Y.-H., Costa, A., Onesti, S., Brunzelle, J. S., Lin, Y., Cann, I. K. O., and Nair, S. K. (2009) *Structure* **17**, 211-222
15. Fletcher, R. J., Bishop, B. E., Leon, R. P., Sclafani, R. A., Ogata, C. M., and Chen, X. S. (2003) *Nature Struct Biol* **10**, 160-167
16. Liu, W., Pucci, B., Rossi, M., Pisani, F. M., and Ladenstein, R. (2008) *Nucleic Acids Res* **36**, 3235-3243
17. Poplawski, A., Grabowski, B., Long, S. E., and Kelman, Z. (2001) *J Biol Chem* **276**, 49371-49377.
18. Kasiviswanathan, R., Shin, J. H., and Kelman, Z. (2006) *J Bacteriol* **188**,

4577-4580

19. Atanassova, N., and Grainge, I. (2008) *Biochemistry* **47**, 13362-13370

20. Sakakibara, N., Kasiviswanathan, R., Melamud, E., Han, M., Schwarz, F. P., and Kelman, Z. (2008) *Nucleic Acids Res* **36**, 1309-1320

21. Barry, E. R., Lovett, J. E., Costa, A., Lea, S. M., and Bell, S. D. (2009) *Proc Natl Acad Sci U S A* **106**, 1051-1056

22. Kasiviswanathan, R., Shin, J. H., Melamud, E., and Kelman, Z. (2004) *J Biol Chem* **279**, 28358-28366

23. Kelman, Z., Lee, J. K., and Hurwitz, J. (1999) *Proc Natl Acad Sci U S A* **96**, 14783-14788

24. Shin, J. H., Jiang, Y., Grabowski, B., Hurwitz, J., and Kelman, Z. (2003) *J Biol Chem* **278**, 49053-49062

25. Heyduk, T., and Lee, J. C. (1990) *Proc Natl Acad Sci U S A* **87**, 1744-1748

26. Licata, V. J., and Wowor, A. J. (2008) *Methods Cell Biol* **84**, 243-262

27. Shin, J. H., Grabowski, B., Kasiviswanathan, R., Bell, S. D., and Kelman, Z. (2003) *J Biol Chem* **278**, 38059-38067

28. Chong, J. P., Hayashi, M. K., Simon, M. N., Xu, R. M., and Stillman, B. (2000) *Proc Natl Acad Sci U S A* **97**, 1530-1535

29. Shechter, D. F., Ying, C. Y., and Gautier, J. (2000) *J Biol Chem* **275**, 15049-15059

30. Moreau, M. J., McGeoch, A. T., Lowe, A. R., Itzhaki, L. S., and Bell, S. D. (2007) *Mol Cell* **28**, 304-314

31. Jenkinson, E. R., and Chong, J. P. (2006) *Proc Natl Acad Sci U S A* **103**, 7613-7618

32. Shin, J.-H., Heo, G.-Y., and Kelman, Z. (2009) *J Biol Chem* **284**, 540-546

33. Rothenberg, E., Trakselis, M. A., Bell, S. D., and Ha, T. (2007) *J Biol Chem* **282**, 34229-34234

34. Lee, J.-K., and Hurwitz, J. (2001) *Proc Natl Acad Sci U S A* **98**, 54-59

35. Ishimi, Y. (1997) *J Biol Chem* **272**, 24508-24513

36. Shin, J.H., Santangelo, T.J., Xie, Y., Reeve, J.N. and Kelman, Z. (2007) *J*

Biol Chem **282**, 4908-4915.

37. A.T. Brunger, P.D. Adams, G.M. Clore, P.Gros, R.W. Grosse-Kunstleve, J.-S. Jiang, J. Kuszewski, N. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson, G.L. Warren, Crystallography & NMR System (CNS), *Acta Cryst*.D**54**, 905-921(1998).

38. A.T. Brunger, Version 1.2 of the Crystallography and NMR System, *Nature Protocols* **2**, 2728-2733 (2007).

Supplementary Table 1. Sequences of substrates used in Processivity and TrpY replacement assays. Shaded boxes in A), B) and C) indicate complimentary sequences; in D) indicates TrpY protein binding sites and all the labeled strands are marked with asterisks.

A) 26 bp duplex (DF61/DF74)

DF61:

*5′TTTGTTTGTTTGTTTGTTTGTTTGTTTGTTTGTTTGCCGACGTGCCAGGCCGACGCGTCCC-3′

DF74:

5′GGGACGCGTCGGCCTGGCACGTCGGCCGCTGCGGCCAGGCACCCGATGGCGTTTGTTTGTTTGTTTGTTTGTTT-3′

B) 55 bp duplex (Trp98/Trp99)

Trp98:

*5’TACATATACACGGAGGTACTATTTAAATCTTGTCTGTATAATAATGTACACTAAGTTT(GTTT)$_{10}$-3’

Trp99:

5’GTCCTACGCTAATCTAAGCCGTTTACTCGCGATTTTGAAAATAGCTTAGTGTACATTATTATACAGACAAGATTTAAATAGTACCTCCGTGTATATGTA-3’

C) 80 bp duplex (PB18/PB19)

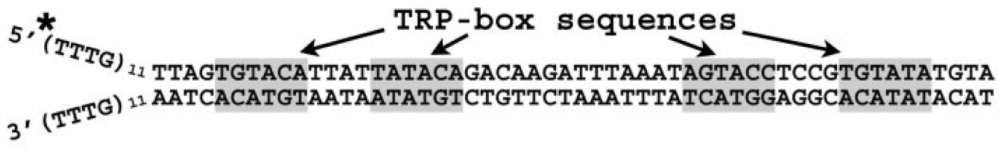
PB18:

*5’TCAACCGTAATGGTGTTGTCCTACGCTAATCTAAGCCGTTTACTCGCGATTTTGAAAATAGCTTAGGTGGAGATCTGATATTTGTTTGTTTGTTTTGTTTG-3’

PB19:

5’TTTGTTTGTTTGTTTGTTTGTATCAGATCTCCACCTAAGCTATTTTCAAAATCGCGAGTAAACGGCTTAGATTAGCGTAGGACAACACCAGTACGGTTGA-3’

D) TrpY binding substrate



CHAPTER 3

STRUCTURAL STUDIES OF N-TERMINAL F172I MUTANT MTMCM USING SMALL-ANGLE X-RAY SCATTERING (SAXS) METHOD

3.1 Introduction

Small-angle scattering of X-rays (SAXS) is a fundamental tool for structural analysis of biological macromolecules in solution, with sizes ranging from a few kilo-Daltons to several mega-Daltons. It does not require special sample preparation, which enables analysis of biological macromolecules and their complexes in nearly physiological environments and direct study of structural responses to changes in external conditions, such as large conformational transitions, assembly or unfolding processes, or intermolecular interactions.

In a scattering experiment, the solution contains macromolecules is exposed to X-rays (wavelength 0.1-0.2 nm), and the scattered intensity, I(s), is recorded as a function of the scattering angle (the angle between the incident and scattered radiation). Scattering intensity from a dilute monodisperse solution of macromolecules (e.g., of purified proteins) is proportional to the spherically averaged single-particle scattering:

$$I(s) = \langle A^2(s) \rangle_\omega \qquad (1)$$

where $s=(s, \Omega)$ is the scattering vector, $s=(4\pi/\lambda)\sin 2\theta$, λ is the wavelength, and 2θ is the scattering angle (1). Only a few particle parameters (radius of gyration Rg, particle volume) are directly obtained from the SAXS data. Additional information about the particle structure is provided by contrast variation (2). The contrast of a particle or its component with a scattering

density distribution $\rho(r)$ in a solvent of density; ρ_s is the average effective density $\Delta\rho=\langle\rho(r)\rangle-\rho_s$. For single-component macromolecules (e.g., proteins), measurements at different ρs allow extraction of the scattering due to the particle shape.

At low angles (resolution of 2 to 3 nm), as x-rays are insensitive to the internal structure, the scattering is essentially determined by the particle shape. Low-resolution shape models were thus constructed on a trial-and-error basis using structural information from other methods. Recently, ab initio modeling approaches have been developed for low-resolution shape restoration representing the shape either by an angular envelope function (3,4), or as an ensemble of densely packed beads (1,5,6), and annealing simulation is employed to fit the scattering data.

In the first general ab initio approach, an angular envelope function of the particle, $r=F(\omega)$, where (r,ω) are spherical coordinates, is described by a series of spherical harmonics. The low-resolution shape is thus defined by a few parameters, the coefficients of this series, that fit the scattering data (2). This approach was further developed and implemented in the computer program SASHA (4). It was demonstrated that, under certain circumstances, a unique envelope can be extracted from the scattering data.

Use of the angular envelope function was, however, limited to relatively simple shapes. A more comprehensive description is achieved in the bead methods (1,5), which use the trial-and-error Debye modelling. A sphere with diameter of maximum particle size, Dmax, which is directly determined from the scattering data, is filled with M densely packed beads. Each bead belongs either to the particle (index=1) or to the solvent (index=0), and the

shape is thus described by a binary string of length M. Starting from a random distribution of 1s and 0s, the model is randomly modified using a Monte Carlo-like search to find a string X to fit the experimental data. As the search models usually contain thousands of beads, the solution must be constrained. In the simulated annealing procedure implemented in the program DAMMIN, an explicit penalty term P (X) is added to the goal function f (X) = χ 2 + P (X) to ensure compactness and connectivity of the resulting shape. Instead of using Debye's formula, the intensity is computed with spherical harmonics to speed up the computation (7).

The assumption of particle homogeneity is a severe limitation of dummy atom models. For single component particles and a single scattering curve, the procedure degenerates to ab initio shape determination as implemented in DAMMIN (2). A more versatile approach to reconstruct protein models from SAXS data has recently been proposed, where the protein is presented by an assembly of dummy residues (DRs) and simulated annealing is used to build a locally 'chain compatible' DR model inside a sphere with a diameter of Dmax. Compared to ab initio shape determination, DR modelling (implemented in the program GASBOR) employs fewer free parameters to account for more experimental information (7).

In spite of the limited resolution and the restricted range of application, a number of successful studies have shown that solution scattering data provide sufficient information to restore particle shapes ab initio.

In this chapter, Small-angle scattering of X-rays (SAXS) was employed to examine mtMCM proteins using the existing N-terminal crystal structure as a priori information. Preliminary structural information obtained by this experiment will be discussed in Section 3.5.

3.2 Materials and Methods

3.2.1 Materials

All mtMCM proteins used in this study were purified as previously described in Chapter II. All mtMCM proteins were further diluted with the same storage buffer (50mM Tris-HCl pH 8.0, 250mM NaCl) for SAXS experiment needs. BSA from Sigma was dissolved in 50 mM Hepes buffer with pH7.5 to 50 mg/ml. Lysozyme was dissolved to 50mg/ml in buffer containing 150mM NaCl, 40mM Na acetate (pH 3.8). To remove high molecular weight protein aggregates from solution, all protein samples were centrifuged at 10 000 rpm for 10 min prior to beam exposure at 4°C. The upper clear liquids were kept for further x-ray scattering experiments.

3.2.2 Small-angle x-ray scattering experiment

SAXS data were collected at 277 K at the Cornell High Energy Synchrotron Source (CHESS) G1 Station. The scattered signals were detected using a Finger lakes CCD detector. At a sample-to-detector distance of both 1430mm and 525mm and an X-ray wavelength of 1.247 Å, the scattering vector s ($s = 4\pi\sin\theta/\lambda$, where 2θ is the scattering angle) ranged from 0.01 to 0.5 $Å^{-1}$, which was calibrated using Silver Behenate as a standard oriented specimen. Protein samples were encapsulated in a 20μl volume sandwiched between two thin parallel mica windows 1 mm apart. This sample cell is designed especially for protein SAXS experiments to reduce radiation damage and sample volume as previously described (8). Protein samples were injected into the sample cell, which was placed in a holder maintained at 4°C during all SAXS measurements. Data collected for

10 s exposures at each spot of a 2D matrix with a spatial interval of 0.6 mm on protein samples to reduce radiation-induced protein damage. Buffer and samples were collected alternatively to reduce error caused by protein aggregations (9). SAXS data were collected at protein concentration 2, 4, 6, 8, 10, 15, 19.3 (mg/ml) for N-terminal mtMCM F172I mutant protein. For each concentration, SAXS data were collected at least twice.

3.2.3 SAXS data analysis

This part of work was done by Xinguo Hong. The overall shapes of the proteins were restored from the experimental data by two independent programs, DAMMIN and GASBOR.

3.3 Preliminary Results and Discussion

The SAXS experiments of mutant N-mtMCM (F172I) protein show that the protein conformation is concentration dependent. At low concentration (~2mg/ml in my experiment), the protein was shown to form dimers in solution (Fig. 3.1).

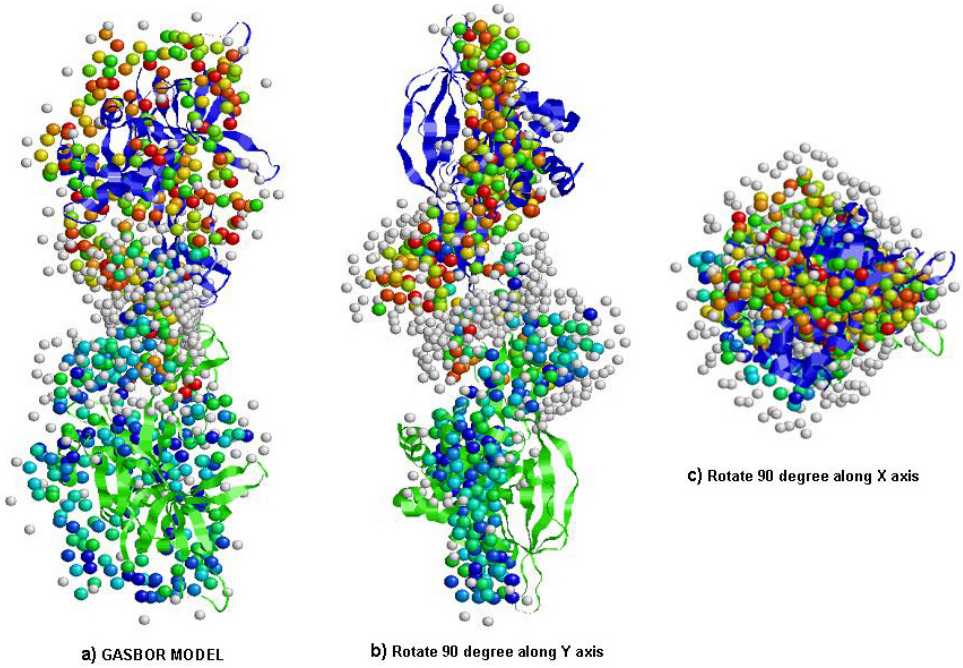


Figure 3.1 GASBOR model of mutant N-terminal mtMCM (F172I) from SAXS data at 2mg/ml protein concentration. Model was generated using GASBOR program fitted by coordinates from published wild-type N-mtMCM protein crystal structure pdb file. Color balls stands for molecular surface of GASBOR model; blue and green ribbons stand for two monomers of N-terminal mtMCM from crystal structure. a) Side view of dimer model; b) Side view rotated 90 degree along Y axis; c) Top view of dimer model.

At concentrations ≥ 4mg/ml. the protein favors a hexameric conformation (Fig. 3.2), but not with a ring structure. The six monomers arrange into a "L" like shape, which is consistent with the N-mtMCM asymmetric unit crystal structure, but definitely different from the reported double-hexamer structure. If this significant conformational change is true, the destabilized MCM4 protein in CHAO3 might has similar changes, and it may give us some ideas about the chromosomal instability caused by this mutation in mice.

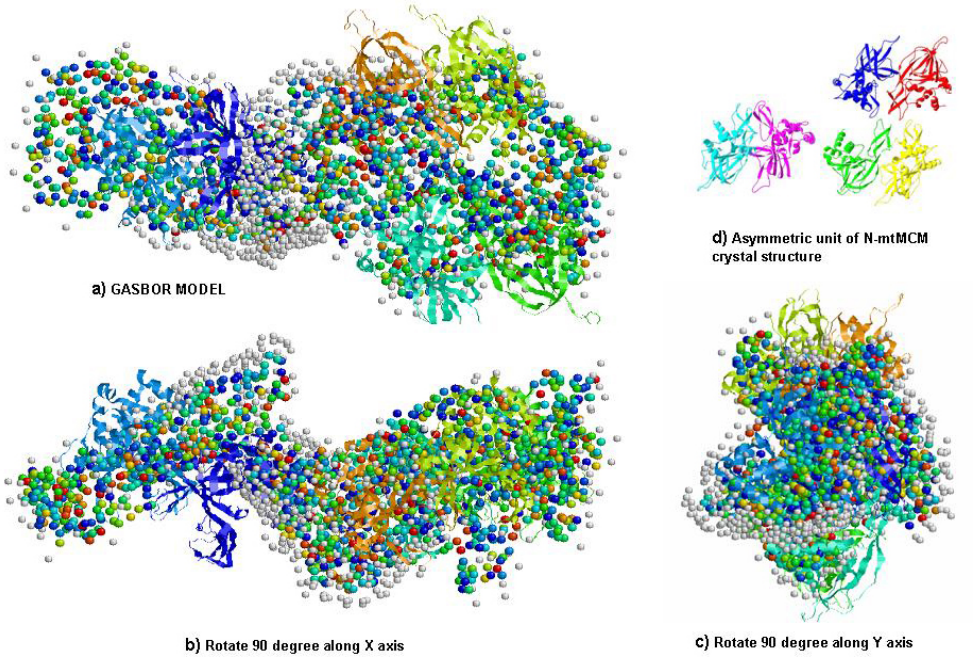


Figure 3.2 GASBOR model of mutant N-terminal mtMCM (F172I) from SAXS data at protein concentrations ≥ 4mg/ml. Model was generated using GASBOR program fitted by coordinates from published wild-type N-mtMCM protein crystal structure pdb file. Color balls stands for molecular surface of GASBOR model; color ribbons stand for N-terminal mtMCM monomers from crystal structure. a) Side view of "L" model; b) Side view rotated 90 degree along Y axis; c) Top view of "L" model.

This result is not consistent with the sizing column and light scattering results (monomer conformation) presented in Chapter II. The possible reason might be that for both sizing column and light scattering experiments, the protein concentrations were less than 2mg/ml. It is possible that when concentration of F172I mtMCM increases, the protein form bigger complex like dimers, hexamers or even larger complex. So at

high concentration, the F172I mtMCM might have multiple conformations in solution and possible equilibrium between all the conformations. This explains why the protein crystallization experiment, which requires high protein concentration and strict homogeneity of protein solution, did not succeed.

The above results give us an idea of overall structure of the N-terminal F172I mutant with the absence of suitable crystals. It might not be very accurate, due to the heavy aggregation under high concentration, but do tell us lots of structural information and conformation possibilities, which will be helpful for further structural or biochemical study on mtMCM proteins. The SAXS experiment itself could be further improved with more data sets.

REFERENCE

1. Svergun D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886.
2. Svergun, D.I., and Koch MH. (2003) Small-angle scattering studies of biological macromolecules in solution. *Rep. Prog. Phys.* **66**, 1735-1782.
3. Svergun, D. I., and H. B. Stuhrmann. (1991) New developments in direct shape determination from small-angle scattering 1. Theory and model calculations. *Acta Crystallogr*. A **47**, 736 –744.
4. Svergun, D. I., V. V. Volkov, M. B. Kozin, and H. B. Stuhrmann. (1996) New developments in direct shape determination from small-angle scattering 2. Uniqueness. *Acta Crystallogr.* A **52**, 419 – 426.
5. Chacon, P., F. Moran, J. F. Diaz, E. Pantos, and J. M. Andreu. (1998) Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm. *Biophys. J.* **74**, 2760 –2775.
6. Feigin, L. A., and Svergun D. I. (1987) Structure Analysis by Small-Angle X-Ray and Neutron Scattering. *New York: Plenum Press.* pp 335.
7. Svergun D.I., and Koch M.H. (2002) Advances in structure analysis using small-angle scattering in solution. *Curr Opin Struct Biol.* Oct; **12**(5), 654-60.
8. Hong X, Hao Q. (2009) Measurements of accurate x-ray scattering data of protein solutions using small stationary sample cells. *Rev Sci Instrum.* Jan; **80**(1), 014303.
9. Hong X, Hao Q. (2009) Combining solution wide-angle X-ray scattering and

crystallography: determination of molecular envelope and heavy-atom sites. *Appl Crystallogr.* Apr 1; **42(**Pt 2), 259-264.

APPENDIXA

ABSTRACTS FOR ADDITIONAL PUBLISHED WORK

Structural basis for modulation of Kv4 K+ channels by auxiliary KChIP subunits. Huayi Wang, Yan Yan, Qun Liu, Yanhua Huang, Yue Shen, Linjie Chen, Yi Chen, **Qiuyue Yang**, Quan Hao, KeWei Wang, and Jijie Chai. *Nat Neurosci.* 2007 January; **10**(1): 32–39. Published online 2006 December 24. doi: 10.1038/nn1822.

Abstract:

KChIPs coassemble with pore-forming Kv4 a subunits to form a native complex in the brain and heart and regulate the expression and gating properties of Kv4 K1 channels, but the mechanisms underlying these processes are unknown. Here we report a co-crystal structure of the complex of human Kv4.3 N-terminus and KChIP1 at a 3.2-A˚ resolution. The structure reveals a unique clamping action of the complex, in which a single KChIP1 molecule, as a monomer, laterally clamps two neighboring Kv4.3 N-termini in a 4:4 manner, forming an octamer. The proximal N-terminal peptide of Kv4.3 is sequestered by its binding to an elongated groove on the surface of KChIP1, which is indispensable for the modulation of Kv4.3 by KChIP1, and the same KChIP1 molecule binds to an adjacent T1 domain to stabilize the tetrameric Kv4.3 channels. Taken together with biochemical and functional data, our findings provide a structural basis for the modulation of Kv4 by KChIPs.